I DEMAND FOR HEALTH AND HEALTH CARE

2 DEMAND FOR HEALTH CARE

Before we left for college, our parents counseled us always to follow the doctor's advice and never to skimp on health care. If the doctor says get a flu shot, get one. If the doctor says get ten flu shots, get all ten – even if they cost $100 each.[1] While our parents' counsel was loving advice, it implies that health care is so valuable that it is worth ignoring any and all economic tradeoffs. In the words of introductory economics, our parents are encouraging us to be *price-inelastic* or *price-insensitive* when it comes to health care.

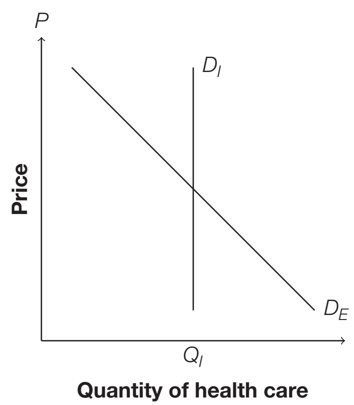


Figure 2.1. *A price-inelastic demand curve,* D_I, *and price-elastic one,* D_E. *This chapter discusses which curve more accurately resembles the demand for health care.*

Are people actually price-insensitive when it comes to health care? Or does demand for health care respond to price, even for health care that may be a matter of life and death?

Figure 2.1 shows two possible demand curves. D_I reflects our parents' advice: the individual with this demand curve is insensitive to price. He wants a certain level of care Q_I and is willing to pay any price to get it. D_E, on the other hand, represents the demand of an individual sensitive to price. She takes price into account when deciding how much care to seek. A non-vertical curve like D_E is said to be *downward-sloping.*

Figure 2.1 may seem simple, but it lies at the center of health economics. Much of the policy debate about how best to organize the provision of health care is grounded in two questions:

- Is the demand curve for health care downward-sloping? Put another way, are people sensitive to the price of health care?
- If the answer to the above question is "yes," people who face different prices or have different willingness to pay get different amounts of care. Do they end up with different health outcomes as a result?

If the answer to the first question is "no," and the demand curve for health care resembles D_I, then the economics of health and medical care is of little interest. The incentives of patients seeking care are inconsequential; instead, there exists a medically optimal level

1 One of the authors, despite being both a professional economist and a medical doctor, gives this questionable advice to his children.

of health (Q_I). Achieving that optimum is a medical problem to be solved by doctors and medical researchers. It is not an economic problem to be solved by utility-maximizing consumers. In this world, health economics is an accounting exercise involving the comparison of different medical treatments and the measurement of different medical outcomes. Health economists studying incentives and markets have little to add.

But the evidence we outline in this chapter overwhelmingly suggests that the answer to the first question above is "yes" – consumers are price-sensitive when it comes to medical care. People with different budget constraints, different life expectancies, different qualities of life evaluate the tradeoff between medical care and other goods differently. One person may decide to skip a knee replacement surgery to pay for his child's tuition. Another person may decide to get laser eye surgery rather than a fancy Christmas gift for his spouse. Determining the right amount of care is not merely a medical matter, but is the outcome of economic tradeoffs that balance the marginal cost of care against the marginal benefit of that care. In other words, demand for health care is downward-sloping.

In many countries, this is rarely an issue because all citizens are entitled to subsidized health insurance or are eligible for free care from the government. But in some countries, notably the US, people must routinely decide how much to pay for care. In those contexts, evidence suggests that people who face different prices or have different abilities to pay for health care receive unequal amounts of health care. But even in countries where patients pay nothing for care at the point of service, whether health care demand is downward-sloping has important consequences for the design of good health care policy, as we will see throughout the book.

Are people who can better afford health care healthier because they receive more and better care? If so, what should a society do, if anything, in response to this possible inequity? These questions underlie the ferocious political debate about health care in every country and motivate much of our study in this textbook.

2.1 Experiments on the demand for health care

Imagine a consultant working on his first case. He is tasked with helping a surgeon predict what will happen to her customer base if she raises prices. To do so, the consultant sets out to plot a demand curve for the surgeon's services.

One method he might use to plot this demand curve is to take a survey of the surgeon's patients and ask them if they would have chosen a different surgeon if the price had been higher or lower. One major problem with this approach is that it ignores the population of people who are not currently patients of the surgeon. A change in price for the surgeon's services may have a different effect on that population. Since the surgeon's patients are likely to be more devoted to her than patients who do not know her, they may be less sensitive to price changes than the people not surveyed.

Alternatively, the consultant could commission a survey of the entire local population. He asks respondents whether they visit surgeons like his client and how much they pay. The main advantage of this approach is that different groups of people – covered by different insurance plans – face different prices for surgical visits. This allows the consultant to construct a demand curve, since he observes different levels of demand at different effective prices. Unlike the first survey, respondents are not asked to conduct any hypothetical thought experiments.

But this survey design is also problematic because the prices that respondents face are not randomly assigned. People choose their insurance plans based on what is advantageous to them. For instance, a respondent who knows he is likely to require surgery will search for an insurance carrier that comprehensively covers surgical services. As a result, people with generous insurance – and therefore facing lower out-of-pocket costs – are exactly the people who are most likely to demand surgery in the first place.

This non-random selection distorts the estimated demand curve because the groups facing each price level differ in important ways. In this case, the people who choose generous insurance are sicker than the typical population, and consequently have higher demand for services. Conversely, people who choose less generous insurance are healthier and have lower demand. Figure 2.2 shows what the measured demand curve D_M might look like if the true demand curve is actually D_T. Under these conditions, the consultant underestimates the demand at the high price P_H and overestimates the demand at the low price P_L.

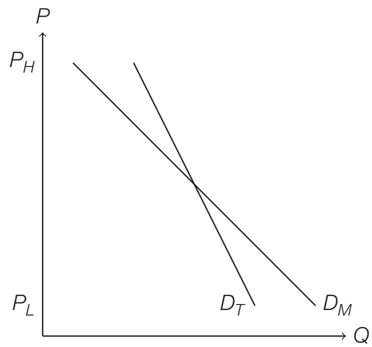


Figure 2.2. *True demand,* D_T, *and measured demand,* D_M, *in a non-randomized study. A non-randomized study such as a broad survey will tend to overestimate health care demand at low prices (and underestimate demand at high prices) because the people who face low prices are the same ones who purchased generous insurance coverage and tend to need more services.*

To calculate a true demand curve, we need to find how the *same* population reacts to different prices. Ideally, we would track the same population in two parallel universes where they face different price levels, but of course this thought experiment cannot be performed in real life. The next best alternative is a **randomized experiment** that assigns treatments randomly to different groups of study participants. Randomization generates experimental groups that are statistically similar. Done correctly, this becomes the best approximation for the parallel universes with actually identical groups. Distortions like the one in Figure 2.2 disappear if there are no meaningful differences between the groups except for the random assignment.

Definition 2.1

Randomized experiment: a study that assigns treatments randomly to different groups of study participants. A **randomized controlled experiment** includes a control group which is randomly chosen and receives either no treatment, a placebo treatment, or the usual treatment they would have received if not enrolled in the trial. Such studies provide the most persuasive evidence on questions of causality in the social sciences and medicine.

Two randomized health insurance experiments

For this chapter, we rely primarily on two influential randomized experiments of health care demand: the **RAND Health Insurance Experiment (HIE)** and the **Oregon Medicaid Experiment**. The RAND study, conducted between 1974 and 1982, was particularly groundbreaking because it was the first large-scale randomized study in which insurance status was randomly assigned, and it is still the only such study ever conducted in the US. Before RAND, there were many non-randomized studies but little consensus about the effects of price on the demand for health care. Since the RAND HIE was published, there has been little dispute that the demand curve for health care is not vertical but in fact downward-sloping.

Parallel universes would be the ideal setting for estimating demand curves, but researchers lack the grant money to build such a testing environment. They rely on randomized experiments instead.
Credit: © rolffimages – Fotolia.com.

For the HIE, the RAND researchers randomly assigned two thousand families from six American cities to one of several different health insurance plans for several years. These plans varied on the generosity of coverage; in particular, the plans had different **copayment rates**.

The copayment rate for an insurance plan is the fraction of the medical bill for which the patient is responsible. Thus, people assigned to different plans had to pay different prices for the same services. There were four different plans: one plan with completely free care (0% copayment rate), and three other **cost-sharing plans** with 25%, 50%, and 95% copayments.[2] Because the plans studied in the RAND HIE differed in only this respect, they are ideal for estimating the effect of price on health care decisions.

Definition 2.2

Copayment rate: the fraction of the medical bill for which the patient is responsible. A **cost-sharing plan** is one with a positive copayment rate, so that costs are shared between the insured and the insurer.

One problem with the RAND HIE is that the health economy has changed in fundamental ways since the 1980s. Consequently, the results found in the RAND HIE may not apply to the demand for health care today. A recent study, the Oregon Medicaid Experiment, provides an interesting update to the RAND results. In general, like the RAND HIE, the Oregon Medicaid Experiment finds downward-sloping demand for health care (Finkelstein et al. 2011).

Unlike the RAND HIE, the Oregon Medicaid Experiment did not assign participants to different insurance plans. Instead, it compared two groups of low-income adult Oregonians: (a) people who won a 2008 lottery to receive the opportunity to apply for public

2 In addition to plans that varied on the copayment rate, there were also other plans including a health maintenance organization (HMO) plan and an individual deductible plan. Interested readers should check out *Free for All?* by Joseph Newhouse (1993) for an in-depth look at the experiment and many details that we lack the space to cover here.

health insurance coverage through Medicaid, and (b) lottery entrants who did not win and were not given a chance to apply for Medicaid. In effect, this lottery randomly assigned insurance coverage to a subset of the winners. Hence, the lottery winners tended to face lower out-of-pocket prices for care.

The approaches of the RAND HIE and the Oregon Medicaid Experiment each have their advantages and disadvantages. The Oregon Medicaid Experiment exclusively focuses on a low-income population, unlike the RAND HIE, which studied a nationally representative population. Furthermore, the RAND HIE used a direct randomization of health insurance coverage, while the Oregon Medicaid Experiment relied on a randomization scheme that was only indirectly related to insurance coverage (Medicaid enrollment was not automatic for lottery winners; they were only 25 percentage points more likely to be covered in the year following the lottery than the lottery losers were). Lastly, the Oregon Medicaid Experiment included an uninsured group that was in part randomly assigned, while the RAND HIE did not include any participants who were totally without insurance.

2.2 Is demand for health care downward-sloping?

If we wish to estimate a demand curve for health care, there are two basic questions to answer before we can even start: How do we define quantity, Q? And how do we define price, P? This may be simple in some cases: in the market for bubblegum, quantity is naturally defined as the number of sticks purchased and price as the cost of a stick.

Matters are more complicated when it comes to health care. A quick visit to the doctor's office is not equivalent to an overnight stay at the hospital. Counting both as one unit of health care is not appropriate, nor is it clear if an overnight stay should count as five doctor's visits or one hundred. Researchers handle this difficulty by measuring separate demand curves for different kinds of care.

Measuring price in health care is also not straightforward. Most health care is paid for by third parties such as private health insurers or the government. Unlike bubblegum buyers, patients pay a *premium*, or an upfront cost, to join an insurance plan, and in exchange they pay lower out-of-pocket prices for each medical service they receive. When calculating demand, the appropriate measure of price is the marginal cost that patients pay when consuming a fixed amount of care. Researchers treat the copayment rate as a measure of price because it is proportional to the marginal cost faced by patients.

The remainder of this section summarizes the experimental evidence on downward-sloping demand for different types of health care.

Outpatient care

If you have ever visited the doctor's office, hospital, or emergency room and gone home the same day, you were the recipient of **outpatient care**. Depending on the severity of your condition, you may not care that your insurance company requires you to pay $20 for the visit. If you have a broken leg, you still want a cast set even if you have to pay the fee. On the other hand, if you just have a runny nose, you might choose chicken soup and *Simpsons* reruns instead of a visit to the doctor.

Definition 2.3

Outpatient care: any interaction with a doctor or other medical care professional that does not involve an overnight stay. Typically, more severe cases will require overnight stays for patient monitoring and recovery, so outpatient cases tend to be less complex.

Outpatient care is also sometimes called **ambulatory care**.

In the health insurance experiments, participants faced different prices for outpatient care. What effect did this have on demand for outpatient services? Both the RAND HIE and the Oregon Medicaid Experiment report the effects of price changes on the demand for outpatient care. The effects are large and show that demand curves for these services are downward-sloping.

Table 2.1. *Evidence for outpatient care: (a) RAND HIE Study. (b) Oregon Medicaid Experiment.*

(a)

	Avg # of annual episodes by condition		
Plan	Total	Acute	Chronic
Free	2.99	2.29	0.70
25%	2.32	1.78	0.54
50%	2.11	1.60	0.51
95%	1.90	1.44	0.46

(b)

	% with visit	No. of visits
Lottery winners	63.6	2.22
Lottery losers	57.4	1.91

Sources: (a) Keeler et al. (1988). With permission from RAND. (b) Amy Finkelstein et al. (2012) The Oregon Health Insurance Experiment: evidence from the first year, *Quarterly Journal of Economics*, 127(3): 1057–1106, Supplementary Data. With permission from Oxford University Press.

Table 2.1(a) shows evidence from the RAND HIE that, as patient cost-sharing increases, the number of episodes of outpatient care decreases sharply. People assigned to the 95% group, for example, had 36% fewer episodes of outpatient care than those in the free plan.

While this result is striking, even more surprising is that patients with *chronic* conditions and *acute* conditions had similar downward-sloping demand. Chronic conditions, such as diabetes and high blood pressure, are health problems that persist over long periods of time and require sustained treatment. Non-chronic or acute conditions are those with sudden onset such as a cold or a broken leg. People assigned to the 95% group had 34% fewer episodes of chronic outpatient care than those in the free plan, and 37% fewer episodes of acute outpatient care.

The Oregon Medicaid Experiment also provides evidence for downward-sloping demand in outpatient care. Table 2.1(b) shows that lottery winners who are more likely to be Medicaid enrollees were 24 percentage points more likely to have an outpatient visit over a six-month period, and had 36% more visits on average, when compared with lottery losers who were unlikely to be on Medicaid.

Inpatient and emergency room care

Imagine that you see your doctor and she tells you that your condition is sufficiently serious that you must stay overnight in the hospital for monitoring. The doctor is admitting you to the hospital for **inpatient care**.

Definition 2.4

Inpatient care: any interaction with a doctor or other medical care professional that involves an overnight stay at a hospital.

Given the severity of your condition, it seems unlikely in this case that you would be too worried about the 20% copayment that the insurance company will charge for the visit, even though the visit will ultimately be very expensive. In fact, even a 50% copayment rate would not deter you from heeding the doctor's orders and staying overnight. What this thought experiment suggests is that demand for inpatient care may not be as sensitive to price as outpatient care is. Does the evidence from these health insurance experiments line up with this intuition?

The evidence on this question is mixed. The data on inpatient care from the Oregon Medicaid Experiment corroborates our intuition in part. Table 2.2(b) shows that though lottery winners had more inpatient visits than lottery losers over a six-month period, the difference was not statistically significant. The evidence does not allow us to conclude that this population is price-sensitive with respect to its demand for inpatient care.

By contrast, the RAND HIE study does find downward-sloping demand for inpatient care. Members of the 95% copayment group were 24% less likely to have inpatient care than members of the free plan in an average year (see Table 2.2(a)). However, the drop in service use at higher prices was smaller than the corresponding drop for outpatient care. Like the Oregon Medicaid Experiment, the RAND HIE finds demand for inpatient care is not as sensitive to price as outpatient care is.

Table 2.2. *Evidence for inpatient care: (a) RAND HIE Study. (b) Oregon Medicaid Experiment.*

(a)

Plan	Avg # of Annual Visits
Free	0.133
25%	0.109
50%	0.099
95%	0.098

(b)

	% with visit	No. of visits
Lottery winners	7.4	0.103
Lottery losers	7.2	0.097

No significant difference at the $p = 10\%$ level.

* Indicates significantly different from the free plan at the $p = 5\%$ level.
** Indicates significantly different from the free plan at the $p = 1\%$ level.

Sources: (a) Keeler et al. (1988). With permission from RAND. (b) Amy Finkelstein et al. (2012) The Oregon Health Insurance Experiment: evidence from the first year, *Quarterly Journal of Economics*, 127(3): 1057–1106, Supplementary Data. With permission from Oxford University Press.

Intuitively, we would expect that the more severe a condition is, the less price-sensitive patients will be with respect to its treatment. This explains why study participants seemed to be more price-sensitive for outpatient care than inpatient care. By this logic, demand for emergency room (ER) care by deathly ill patients should be completely price-insensitive: regardless of costs, people will seek ER care in cases of life and death.

Nonetheless, there is evidence that demand for emergency care slopes downward. While the Oregon Medicaid Experiment finds no statistically significant difference in the rates of ER visits (Table 2.3(b)), participants in the RAND HIE were sensitive to price.

Table 2.3. *Evidence for emergency care: (a) RAND HIE Study. (b) Oregon Medicaid Experiment.*

(a)	
Plan	Probability of ER use
Free	22%
25%	19%*
50%	20%
95%	15%**

(b)		
	% with visit	No. of visits
Lottery winners	26.7	0.48
Lottery losers	26.1	0.47

No significant difference at the $p = 10\%$ level.

* Indicates significantly different from the free plan at the $p = 5\%$ level.
** Indicates significantly different from the free plan at the $p = 1\%$ level.

Sources: (a) Newhouse (1993). With permission from RAND. (b) Amy Finkelstein et al. (2012) The Oregon Health Insurance Experiment: evidence from the first year, *Quarterly Journal of Economics*, 127(3): 1057–1106, Supplementary Data. With permission from Oxford University Press.

Table 2.3(a) shows that people in the cost-sharing groups were less likely to visit the ER than people on the free plan.

This result is at least a bit surprising; if all ER visits are truly for dire emergencies, this means that demand for life-saving care is downward-sloping. But not every ER visit is actually a matter of life or death. Many patients visit the ER for less urgent matters because they do not have a regular primary care doctor, have nowhere else to go, or have overestimated the severity of their condition (Garcia et al. 2010). When patients face high prices, it is these nonurgent ER visits that are most likely deterred.

Other evidence on outpatient, inpatient, and emergency care

Other studies also provide convincing evidence that demand for inpatient and emergency care slopes downward. The China Rural Health Insurance Experiment (CRHIE), conducted by RAND researchers and largely modeled on the RAND HIE, finds similar patterns. Like the RAND experiment, the CRHIE randomly assigned participants to insurance plans with different copayment rates. The participants were drawn from 26 villages in rural China, and their health expenditures were tracked over a two-year period from 1988 to 1989 (Cretin et al. 2006).

CRHIE participants with more generous insurance plans incurred more medical costs than participants with less generous plans, just as in the RAND HIE and the Oregon Medicaid Experiment (Table 2.4). As with the American experiments, the difference is more pronounced for outpatient expenditures: those in the most generous plan incurred more than twice as much outpatient cost per person compared with the people in the least

Table 2.4. *Mean health care expenditures in the CRHIE, in yuan.*

Copayment rate	Outpatient care	Inpatient care
20%	13.49	4.51
30%	12.04	4.18
40%	10.72	3.88
50%	9.52	3.61
60%	8.42	3.36
70%	7.43	3.14
80%	6.54	2.95

Source: Based on Table 8 in Cretin et al. (2006). With permission from RAND.

generous plan, but only about 53% more inpatient costs. Just like RAND HIE participants and Oregon Medicaid lottery entrants, health care consumers in rural China show signs of price-sensitivity, especially when it comes to outpatient care.

Data from non-experimental studies provides more corroborating evidence for both downward-sloping demand and different rates of price-sensitivity for different kinds of care. One non-experimental strategy researchers have used is to study people who suddenly change insurance status from uninsured to insured. If demand is indeed downward-sloping, we would expect to see an increase in the use of medical services when an individual becomes insured. But if urgent care is less price-sensitive, then we would expect to see less of a jump in those services when individuals gain insurance coverage.

Most US citizens become eligible to enroll in Medicare when they turn 65, including many people who were uninsured before that age. Card et al. (2009) use this fact in an analysis of California hospital admissions data to measure the effect of insurance coverage on demand. The number of planned hospital admissions per patient jumps by 15% from age 64 to age 65 when Medicare coverage begins. Even if this jump reflects "pent-up demand" – a phenomenon where individuals who know they will soon have insurance access delay costly procedures – it is evidence that people are sensitive to price when it comes to planned hospital admissions. By contrast, unplanned hospitalizations that begin in the ER increase by only 2.5% at that age (Figure 2.3). Again, when life is at stake, the price of care seems to matter a lot less, or not at all.

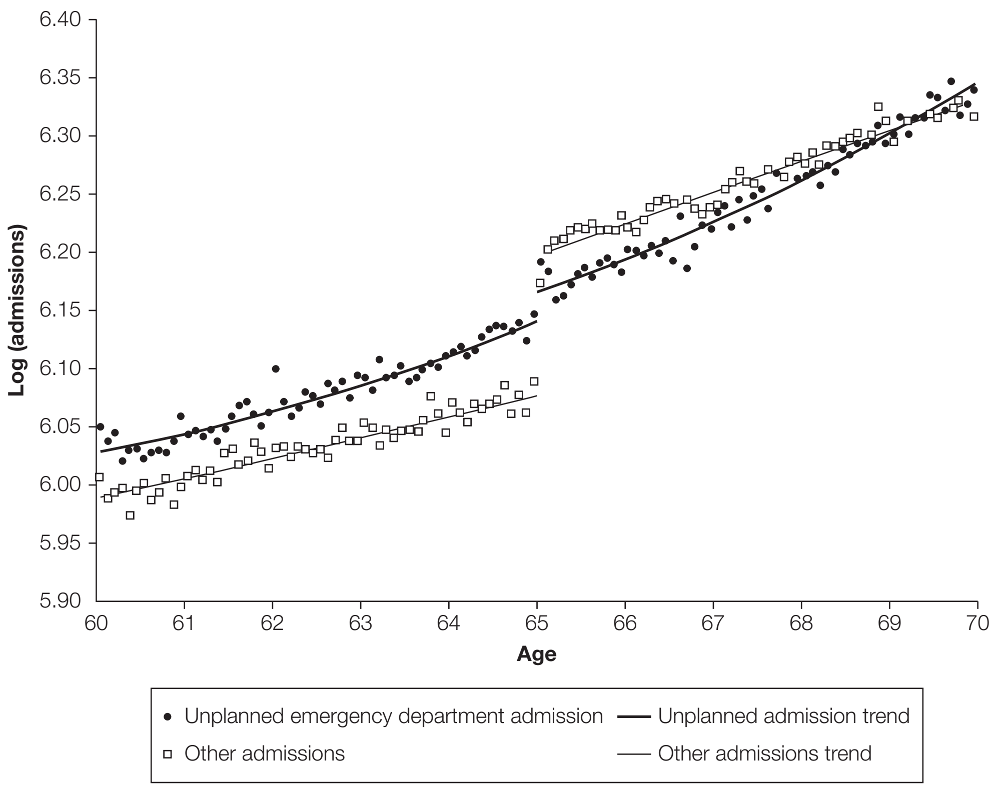


Figure 2.3. *Emergency and non-emergency visits by age.*

Source: Figure 2 in D. Card, C. Dobkin and N. Maestas (2009). Does Medicare save lives?, *Quarterly Journal of Economics*, 124(2): 597–636, by permission of Oxford University Press.

A related approach takes advantage of a change in policy extending insurance to a well-defined group in the population. For instance, in 2000, the French government made a free complementary insurance plan available to the poorest 10% of French residents. In an analysis of this policy, Grignon and Perronnin (2008) find that people in this group who previously lacked complementary coverage increased their use of medical care services.

Pediatric care

This child probably wishes his parents were more *price-sensitive.*
Credit: Bananastock.

We have seen plenty of evidence that demand slopes downward for all different sorts of care. People trade off their health against other goals in almost every conceivable situation. But consider now pediatric care, that is, care for infants and children which is typically paid for by parents. Despite all the evidence we have seen for downward-sloping demand, it is hard to imagine that parents would skimp on care for their children because of price.

And yet evidence from the RAND HIE shows parents are price-sensitive even with respect to health care for their children. Families on the free plan were significantly more likely to have sought immunizations and other preventative care for infants, toddlers, and young children (ages 0–6) than families with positive copayment rates (Table 2.5). This is evidence that even the demand curve for pediatric care slopes downward, although this pattern does not seem to extend to older children and adolescents (ages 7–16).

Table 2.5. *Percentage with preventative pediatric care over three years, by age and care type.*

	0–6 years		7–16 years	
	Immunization	Any preventative	Immunization	Any preventative
Free	58.9	82.5	21.2	64.8
Copayment	48.7*	73.7*	21.7	59.6

* Statistically significant discrepancy from free plan.

Source: Newhouse (1993). With permission from RAND.

Other types of care

The RAND HIE and Oregon Medicaid Experiment also gathered use and spending data for **mental health care**, **dental care**, and **prescription drug use**. In each case, both studies find strong evidence of downward-sloping demand.

Table 2.6 displays evidence that per-capita expenditures on ambulatory mental health care depend on plan type. Participants on the free plan used more than twice as much mental health care in dollar terms as the participants with the 95% copayments.

The demand for dental care is even more price-sensitive than the demand for outpatient medical care. Nearly 58% of low-income participants on the free plan visited the dentist annually, compared with only 40% of the low-income participants on the 95% plan, and incurred about 47% more total dental expenditures (Table 2.7). Similar patterns hold for

Table 2.6. *Per-capita mental health expenditures, by plan type.*

Plan	Mean expense ($)	Percentage of free plan
Free	42.2	–
25%	28.4	67%
50%	13.1	33%
95%	18.1	43%

Source: Newhouse (1993). With permission from RAND.

Table 2.7. *Dental care utilization by income level.*

	Low-income group†		High-income group†	
	Percentage with any use	Average expenditures ($)	Percentage with any use	Average expenditures ($)
Free	57.8	317	74.7	339
95%	39.8*	216*	61.3*	234*

* Statistically significant discrepancy from free plan.

† The low-income group comprises the third of households with the lowest incomes. The high-income group comprises the third of households with the highest incomes.

Source: Newhouse (1993). With permission from RAND.

high-income participants, indicating that families on tight budgets are not the only ones sensitive to price when taking care of their teeth.

Prescription drugs are no exception to the general rule of downward-sloping demand. Both the RAND HIE and the Oregon study find evidence that patients facing lower prices use more prescription drugs. The case of antibiotic drugs is particularly interesting. When patients have an illness caused by a bacterial condition, it is medically appropriate to prescribe antibiotics. But when a patient suffers from a viral infection, antibiotics are powerless against the patient's condition and even potentially harmful because antibiotic use can breed resistant bacterial strains.

Despite the futility of antibiotics for treating viruses, patients with colds (which are caused by viruses) often pressure their doctors to prescribe antibiotics anyway, not realizing the uselessness of such treatment. Doctors know better, but in busy practices they are prone to relent and prescribe antibiotics anyway. The RAND HIE provides evidence of this fact: patients with viral infections commonly received antibiotic prescriptions (Table 2.8).

Raising the price of antibiotics has both a bad effect and a good effect. Among patients with bacterial illnesses, those in the free plan received more antibiotic prescriptions than those in the cost-sharing plans. Thus, it is possible that some sick patients in the cost-sharing plans went without antibiotics, when such prescriptions would have been useful. Among patients suffering viral infections, cost-sharing also reduced the use of antibiotics.

Table 2.8. *Antibiotic use in the RAND HIE.*

	No. of antibiotics per person	
Plan	Bacterial conditions	Viral conditions
Free	0.47	0.17
Copay	0.24**	0.08**

** Statistically significant discrepancy from the free plan.

Source: Keeler et al. (1988). With permission from RAND.

In this case, the lower price of antibiotics for people on the free plan offered no health benefits, while potentially abetting the evolution of resistant bacterial strains.

2.3 Measuring price sensitivity with elasticities

Evidence from the randomized experiments establishes that the demand for inpatient and outpatient care slopes downward, but we have not directly examined the demand curve implied by this evidence. Figure 2.4 plots data on the use of outpatient care and dental care from Keeler et al. (1988) in the form of a traditional demand curve.

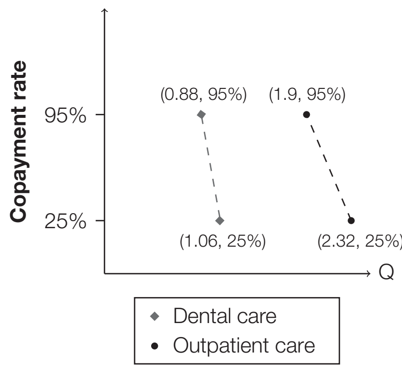


Figure 2.4. *Data on outpatient and dental care.*

Source: Keeler et al. (1988). With permission from RAND.

One simple measure of price sensitivity for each type of care is the slope of the line plotted between the two points of measured demand. The problem is that the units of the quantity demanded are not comparable. A dental visit is not the same as an outpatient visit. Hence, the fact that one slope is greater than the other is not meaningful for comparing price sensitivities between the two goods. Instead, we need a measure of price sensitivity that is not affected by the units in which either price or quantity are measured. The **elasticity of demand** provides just such a measure.

Definition	**2.5**

Elasticity of demand: the ratio that represents how a fixed percentage change in the price of a good leads to a change in the quantity demanded, measured as a percentage change from the original quantity.

Let Q_1 be the original quantity demanded at the price P_1 and let Q_2 be the new quantity demanded after the price changes from P_1 to P_2. The **elasticity** ϵ between these two points is defined as

$$\epsilon = \frac{(Q_2 - Q_1)/Q_1}{(P_2 - P_1)/P_1} \tag{2.1}$$

For instance, suppose an individual starts with an insurance plan with a 25% copayment rate and switches to a plan with a 95% copayment rate. This represents a 280% increase in the price of care for the individual: $^{(95-25)}/_{25} \times 100\% = 280\%$. Figure 2.4 shows how her quantity of outpatient care demanded changes with the switch in insurance: it decreases from 2.32 episodes per year to 1.9, an 18% decline. The elasticity of demand for this individual is

$$\epsilon = \frac{(Q_2 - Q_1)/Q_1}{(P_2 - P_1)/P_1} = \frac{(1.9 - 2.32)/2.32}{(95\% - 25\%)/25\%} = \frac{-0.18}{2.8} = -0.06$$

A similar calculation shows that the elasticity of demand for dental care is also −0.06, even though the slope of the demand curve for dental care looks steeper; comparing slopes can be deceptive. Elasticity is useful for comparing demand curves for various goods, or the demand curve for the same good in different places or settings, because it is *unitless*. We can use elasticities to compare downward-sloping demand for different types of health care or for the same type of care in different studies. While it was already evident that the demand curves for these goods are downward-sloping, these calculations affirm that demand for these goods is relatively inelastic ($-1 < \epsilon < 0$).

One problem with the definition of elasticity ϵ we have used so far is it treats price changes from P_1 to P_2 and P_2 to P_1 asymmetrically. For instance, if the individual had instead switched from a 95% copayment to a 25% copayment plan, her elasticity of demand for outpatient care would be

$$\epsilon = \frac{(Q_2 - Q_1)/Q_1}{(P_2 - P_1)/P_1} = \frac{(2.32 - 1.9)/1.9}{(25\% - 95\%)/95\%} = \frac{0.22}{-0.74} = -0.30$$

Analogously, her elasticity of demand for dental care would be −0.28. The choice of a starting point for price makes a big difference in the ultimate calculation of elasticity.

We would prefer a measure of elasticity that does not require us to pick a starting point and treats price increases and decreases symmetrically. One way around this problem is to measure elasticity at the midpoint between the two endpoints of the demand curve, rather than at the endpoints themselves. This alternate formulation is called the **arc elasticity**.

Definition 2.6

Let (Q_1, P_1) and (Q_2, P_2) be two points on a single demand curve. The **arc elasticity** ϵ_{arc} between these two points is defined as

$$\epsilon_{arc} = \frac{\Delta Q/(Q_1 + Q_2)}{\Delta P/(P_1 + P_2)} \tag{2.2}$$

where $\Delta Q = Q_2 - Q_1$ and $\Delta P = P_2 - P_1$.

When we apply formula (2.2) to the data from Figure 2.4, we find that the arc elasticity of demand for outpatient care is −0.17 and for dental is −0.16.

Figure 2.5 shows the arc elasticity of demand for various health care goods implied by the RAND HIE, alongside estimated elasticities for several other common goods. Demand for medical care goods tends to be inelastic ($-1 < \epsilon_{arc} < 0$). Meanwhile, the demand for goods such as restaurant meals and fresh tomatoes is more elastic ($\epsilon < -1$).

Evidence from the RAND HIE and the Oregon Medicaid Experiment suggests that demand for medical care is quite inelastic. But an arc elasticity calculation is an average measure, so it may not fully capture the range of behavior of all health care consumers. For instance, it is possible that those who spend the most on health care are less sensitive to the price of additional care. Kowalski (2009) analyzes non-experimental data from a large US employer and finds that the 5% of the population that spent the most on medical care are far less price-sensitive than the general population, while typical members of the population are fairly *elastic* in their demand.

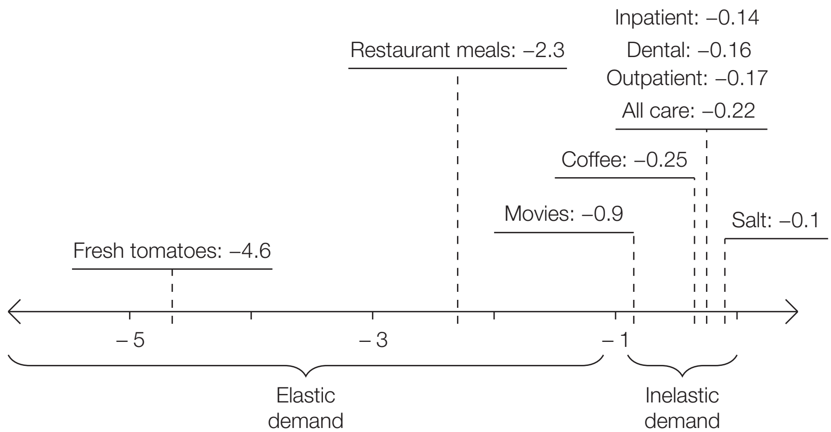


Figure 2.5. *Elasticities of various goods.*

Source: Developed from Newhouse (1993) and Gwartney et al. (2008).

2.4 Does the price of health care affect health?

The previous section showcased the overwhelming evidence that demand for health care is downward-sloping. Though more exigent needs such as inpatient care and ER care seem less sensitive to price, it is clear that people do sometimes skimp on health care, despite parents' advice. The next important question is: How much impact, if any, does the price of health care have on *health* itself?

The simplest way to compare the health of different groups is with mortality rates. The Oregon Medicaid Experiment tracked participant survival rates, and did not find a significant difference between the survival of Medicaid lottery winners and lottery losers in the first year after the lottery (Table 2.9(b)). By contrast, a different study of the Medicaid program in the US does find that insurance coverage reduces mortality. Meyer and Wherry (2012) analyze the expansion of health insurance to a population of poor children in the US. They find that providing Medicaid to black children increased their probability of survival to their late teens by between 13% and 18% relative to similar black children who were not eligible for the Medicaid expansion because they were born before an arbitrary age cutoff date.

Like the Oregon Medicaid Experiment, the RAND HIE also failed to find a mortality difference between treatment groups. The risk of dying for all participants is nearly

Table 2.9. *Evidence on mortality rates: (a) RAND HIE Study. (b) Oregon Medicaid Experiment.*

(a) Relative mortality rate		
	All participants	High-risk participants[a]
Free	0.99	1.90
Copay	1	2.10*

(b) Absolute two-year mortality rate	
	All participants
Lottery winners	0.8%
Lottery losers	0.8%

[a] Participants were classified as high risk based on their blood pressure, cholesterol levels, and smoking habits at the beginning of the study.

* Indicates significantly different from the free plan at the $p = 5\%$ level.

Sources: (a) Newhouse (1993). With permission from RAND. (b) Amy Finkelstein et al. (2012) The Oregon Health Insurance Experiment: evidence from the first year, *Quarterly Journal of Economics*, 127(3): 1057–1106, Supplementary Data. With permission from Oxford University Press.

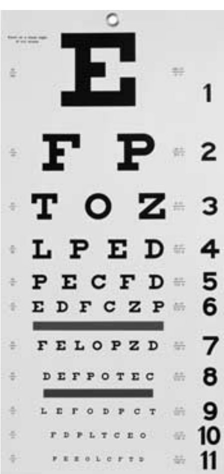
Table 2.10. *Health indicators by insurance plan in the RAND HIE.*

Condition	Free plan	Copay plan
FEV_1^a	95.0	94.8
Diastolic blood pressure (mm Hg)	78.0	78.8*
Cholesterol (mg/dl)	203	202
Glucose (mg/dl)	94.7	94.2
Abnormal thyroid level (% of sample)	2.4	1.7
Hemoglobin (g/100 ml)	14.5	14.5
Functional far vision (Snellen lines)	2.4	2.5*
Functional near vision (Snellen lines)	2.35	2.44*
Chronic joint symptoms (% of sample)	30.0	31.6

[a] FEV is forced expiratory volume in 1 second.
* Indicates significantly different from the free plan at the $p = 5\%$ level.
Source: Newhouse (1993). With permission from RAND.

identical under the free plan and under the copayment plans: those in the free plan were 99% as likely to die during the experiment as those in the cost-sharing plans (Table 2.9(a)).

But free insurance did seem to have an impact on the mortality rates among the most vulnerable participants in the RAND HIE. At the beginning of the study, participants were categorized into risk categories based on their blood pressure, cholesterol level, and smoking habits. The high-risk people on the free plan were 10% less likely to die than high-risk participants on the cost-sharing plans.



A Snellen chart for testing vision. According to the RAND HIE, members on the free plan could read an average of 0.1 lines farther down the chart than members on the cost-sharing plan. Credit: Getty.

Mortality is an extreme outcome. So the fact that people in different plans did not differ in their mortality rates does not mean that more affordable care does not yield any health benefits. In addition to mortality, the RAND HIE tracked various health indicators such as cholesterol levels and blood pressure. Table 2.10 shows some of the differences in health indicators between members of the free plan and members of cost-sharing plans measured in the RAND HIE. Out of 23 health comparisons, the only statistically significant differences between the plans were in blood pressure, myopia, and presbyopia.[3] With the exception of blood pressure, there were no medically important differences as a result of more generous insurance coverage. Even this difference in blood pressure was concentrated among the low-income and high-risk study participants (Newhouse 1993).

As in the RAND HIE, investigators in the Oregon Medicaid Experiment measured the effect of insurance coverage on health status. One key measure evaluated was self-reported health. By this measure, Medicaid lottery winners were healthier than lottery losers twelve months after the lottery. Lottery winners reported better overall health, more healthy days, and lower rates of depression than lottery losers did (Table 2.11). This survey evidence suggests that the lottery winners benefited as a result of the lower price of health care they faced, even though Table 2.9(b) shows there were no mortality differences.

The Oregon investigators found that lottery winners with diabetes had lower blood sugar levels, and that lottery winners with depression were more likely to be receiving

3 Both myopia and presbyopia are vision-related problems. Myopia is near-sightedness, and presbyopia is far-sightedness.

Table 2.11. *Effect of lottery win on health in the first year of the Oregon Medicaid Experiment.*

	Lottery winners	Lottery losers
Survived one year after lottery	99.2%	99.2%
Self-reported health good	58.7%	54.8%**
Self-reported health fair or better	88.9%	86.0%**
Health about the same or better over last six months	74.7%	71.4%**
No. of days physical health good	22.2	21.9*
No. of days mental health good	19.3	18.7**
Did not screen positive for depression in last two weeks	69.4%	67.1%**

* Difference between lottery winners statistically significant at the $p = 5\%$ level.
** Difference between lottery winners statistically significant at the $p = 1\%$ level.

Source: Amy Finkelstein et al. (2012) The Oregon Health Insurance Experiment: evidence from the first year, *Quarterly Journal of Economics*, 127(3): 1057–1106, Supplementary Data. With permission from Oxford University Press.

treatment. However, as in the RAND study, Oregon investigators found little statistically significant evidence of health differences for many health outcomes between lottery winners and lottery losers in the first two years of the experiment (Baicker et al., 2013). Rates of hypertension and high cholesterol, for instance, were statistically quite similar. Together, the studies imply that more affordable care does not measurably improve health for most of the population, but can improve health for at-risk groups.

Findings from several non-experimental studies largely support this conclusion. Studies of vulnerable sub populations, including Wisconsin car accident victims, low-income pregnant mothers on Medicaid, elderly Seattle-area veterans, and dying HIV patients, all find that more affordable health care improves health (Levy and Meltzer 2004; Doyle 2005; Currie and Gruber 1996; Fihn and Wicher 1988; Goldman et al. 2001). On the other hand, evidence on broader populations of self-employed workers finds no mortality differences as a result of insurance coverage, just like the RAND HIE did (Perry and Rosen 2004).

2.5 Conclusion

It seems clear that demand for health care is in fact downward-sloping. Evidence shows unequivocally that people take price into account when deciding how much medical care to seek, even for serious conditions. In short, economic tradeoffs matter even in the world of health. This means that economic analysis is relevant in the world of health care.

Downward-sloping demand implies a fundamental tradeoff for the design of any health care system. If health care is free at the point of service (as it is in many countries), people will demand a lot of care, even care that is not particularly useful for improving their health. For instance, in the RAND HIE, members of the free plan had the same health status outcomes and mortality rates on average as the people on the cost-sharing plans.

On the other hand, if health care is not free, some patients will respond by seeking less care. High prices for health care may have deadly consequences for the most vulnerable segments of the population such as the poor and the chronically ill. Ultimately, this tradeoff between economic efficiency and social equity forms the core of the health policy debate.

2.6 Exercises

Comprehension questions

Indicate whether the statement is true or false, and justify your answer. Be sure to cite evidence from the chapter and state any additional assumptions you may need.

1 Unlike with most types of goods, deriving a demand curve for health care is quite simple because people rarely skimp on health care.

2 The RAND study was especially useful for measuring price elasticities because it randomly assigned insurance plans to participants (as opposed to letting them choose).

3 The Oregon Medicaid Experiment is not truly "randomized" because lottery winners did not all end up with insurance, and some lottery losers did end up with insurance.

4 The RAND HIE found that people assigned to the free health plan had the same rate of hospitalization as people assigned to the cost-sharing plans.

5 In the RAND HIE, the arc elasticity of demand for inpatient care was larger (in absolute value) than the arc elasticity of demand for outpatient care.

6 Unlike the usual measure of elasticity, an arc elasticity can be calculated from just one price–quantity data point.

7 Both the RAND and Oregon studies find that demand for health care is approximately unit elastic, that is, $\epsilon \approx -1$.

8 In the RAND HIE, being assigned more generous insurance did not generally improve participants' health outcomes, except among certain subgroups.

9 To date, no major health insurance experiment has studied the impact of *un*insurance, just different levels of insurance.

10 Results from the Oregon Medicaid Experiment suggest that having health insurance has a positive impact on health status.

Analytical problems

11 Suppose you are collecting data from a country like Japan where the government sets the price of health care. Each prefecture in Japan has a different set of prices (for example, Tokyo has higher prices than rural Hokkaido). Data for 1999 is displayed in Table 2.12.

Table 2.12. *Outpatient utilization in Tokyo and Hokkaido, 1999.*

Region	Outpatient visits	Price/visit
Tokyo	1.25/month	20¥
Hokkaido	1.5/month	10¥

a What is the arc price elasticity of demand for health care consumers in Japan (using only this data)?

b Suppose that incomes are generally much higher in Tokyo than Hokkaido. Is your answer to the last question an overestimate or underestimate of price elasticity? Justify your answer. [*Hint*: It may be helpful to plot the data points from Table 2.12 and consider likely demand curves for Tokyo and Hokkaido.]

c Using your estimated elasticity, what would the demand for health care be if the price in Tokyo were raised to 30¥ per visit? What would the demand in Hokkaido be if the price were lowered to 5¥ per visit?

You continue your observations of the Japanese health care system into the year 2000. For inscrutable reasons having to do with internal Japanese politics, the government changed the price in both Tokyo and Hokkaido that year, and you observe the demand recorded in Table 2.13.

Table 2.13. *Outpatient utilization in Tokyo and Hokkaido, 2000.*

Region	Outpatient visits	Price/visit
Tokyo	1.0/month	30¥
Hokkaido	1.2/month	15¥

d Calculate the price elasticity of demand for health care in Japan using only data from the year 2000.

e Use data from both years to calculate the elasticity of demand for health care for Tokyo and Hokkaido separately.

f Using your estimated elasticities, what would the demand for health care in each prefecture be if the price were raised to 60¥ per visit next year (for both prefectures)?

g Combine the Tokyo and Hokkaido estimates from Exercise 11(e) to get a single estimate of the price elasticity of health care demand for all of Japan. Assume that Tokyo is five times as populous as all of Hokkaido.

12 **Preventative care** refers to care taken to prevent future diseases rather than to treat current ones. Compared with ER care, preventative care is rarely urgent, and benefits can be difficult to measure: if you had the flu vaccine this year but did not catch the flu, it is impossible to tell if it was the shot or assiduous hand-washing that preserved you.

a Given this description of preventative care, would you expect preventative care to be more or less price-sensitive compared with inpatient care? Why?

b Table 2.14 shows evidence on preventative care from the RAND HIE. Summarize the data in the table and note any interesting patterns. Was your prediction correct?

Table 2.14. *Percentage with preventative care in the three years from the RAND HIE Study.*

	Males 17–44	Males 45–64	Females 17–44		Females 45–64	
	Any care	Any care	Any care	Pap test	Any care	Pap test
Free	27.2%	39.1%	83.7%	72.2%	76.9%	65.0%
Copay	23.1%	27.4%	76.9%**	65.8%	65.3%**	52.8%**

** indicates statistically significant difference from the free plan at the $p = 1\%$ level.

Source: Newhouse (1993). With permission from RAND.

13 In this exercise, assume that the term “admission” in Table 2.15 refers to inpatient care, while “any use” refers to inpatient and outpatient care.

Table 2.15 contains a lot of information. Without looking at any specific values, summarize what type of data the table contains. Give an example of a broad question about income levels and demand for health care that the table might have the potential to answer.

Table 2.15. *Various measures of predicted annual use of medical services by income group.*

	Income			Significance tests t on contrast of:	
Plan	Lowest third mean	Middle third mean	Highest third mean	Middle vs. lowest thirds[a]	Highest vs. lowest thirds[a]
Likelihood of any use (percent)					
Free	82.8	87.4	90.1	4.91	5.90
Family pay					
25 percent	71.8	80.1	84.8	5.45	6.28
50 percent	64.7	76.2	82.3	4.35	4.86
95 percent	61.7	68.9	73.8	3.96	4.64
Likelihood of one or more admissions (percent)					
Free	10.63	10.14	10.35	−0.91	−0.35
Family pay					
25 percent	10.03	8.44	7.97	−2.95	−2.75
50 percent	9.08	8.06	7.77	−1.78	−1.66
95 percent	8.77	7.38	7.07	−2.79	−2.46
Expenses (1984 $)					
Free	788	736	809	−1.78	0.53
Family pay					
25 percent	680	588	623	−3.17	−1.47
50 percent	610	550	590	−1.89	−0.49
95 percent	581	494	527	−3.09	−1.41

Note: Excludes dental and outpatient psychotherapy. Predictions for enrollment population carried forward for all years of the study.

[a] The t-statistics are corrected for intertemporal and intrafamily correlation. The statistics test the null hypothesis that the mean of the middle (highest) third equals the mean of the lowest third; for example, the 4.91 figure implies we can reject at the 0.001 level the hypothesis that in the free plan the likelihood of any use for the lowest and middle thirds of the income distribution are equal.

Source: Manning et al. (1987). Reprinted with permission from the American Economic Association.

Essay questions

14 Here is a selection from an abstract of a recent study entitled "The effect of health insurance coverage on the use of medical services" by Michael Anderson, Carlos Dobkin, and Tal Gross (2010). NBER Working paper No. 15823.

> Substantial uncertainty exists regarding the causal effect of health insurance on the utilization of care. Most studies cannot determine whether the large differences in healthcare utilization between the insured and the uninsured are due to insurance status or to other unobserved differences between the two groups. In this paper, we exploit a sharp change in insurance coverage rates that results from young adults "aging out" of their parents insurance plans to

estimate the effect of insurance coverage on the utilization of emergency department (ED) and inpatient services. [*In the US, children are eligible for insurance coverage through their parents' insurance only up to their 23rd birthday, at which point they lose eligibility.*] Using the National Health Interview Survey (NHIS) and a census of emergency department records and hospital discharge records from seven states, we find that aging out results in an abrupt 5 to 8 percentage point reduction in the probability of having health insurance. We find that not having insurance leads to a 40 percent reduction in ED visits and a 61 percent reduction in inpatient hospital admissions.

a What two groups are being compared in this study?

b Identify at least one important methodological difference between the design of this study and the RAND HIE. Give a hypothetical reason why this difference would bias the results.

c Are the findings of this study generally consistent with the findings from the Oregon Medicaid Experiment?

Students can find answers to the comprehension questions and lecturers can access an Instructor Manual with guideline answers to the analytical problems and essay questions at **www.palgrave.com/economics/bht**.

3 DEMAND FOR HEALTH: THE GROSSMAN MODEL

Is health something that happens to you, or something that you choose? Clearly, it is a little of both. A heart attack is an unpredictable event that can happen at any time, even to the young and fit. Getting hit by a bus is also bad for your health, and may be outside your control as well. But there are many actions you can take that reduce the likelihood of heart attacks or bus accidents. Cutting back on Big Macs, for instance, might reduce your heart attack risk, while cloistering yourself in your home virtually eliminates the prospect of a bus collision.

When Sid Meier's Civilization *was first released in 1991, one of the authors spent sixteen straight hours crushing the Mongols. These were hours he could have spent jogging instead.*
Credit: Edward Mallia – iStockphoto.com.

The situation is even more complicated when we consider that a single, seemingly beneficial action can have a wide variety of costs and consequences. For instance, you can take up jogging to keep in shape and avoid heart attacks, but this may increase your exposure to buses. And even if you manage to steer clear of buses, taking time out of your day to jog will reduce the time available for other healthy activities like body-building or pilates. As always, every use of time has an opportunity cost and every decision implies a tradeoff.

Furthermore, health is not the only important thing in life. You probably value many other things, not all of which will extend your lifespan. Maybe, like the authors, you enjoy the occasional jigsaw puzzle or video game. These pursuits, while admirable, reduce the time available for activities that improve health. Maybe you also enjoy snacking on chocolate-chip cookies, which is not very time-consuming but may make you gain weight.

Finally, your health decisions need to be considered in the context of your entire lifespan. Health is a form of capital – it is a valuable asset that pays dividends throughout your life but depreciates as you age. So managing your health over your lifetime is an economic problem that is similar, in some ways, to managing a stock portfolio.

We need a framework that captures all of the complex tradeoffs involved in health management. It should model health as a consumption good, an input into the enjoyment of other goods, and a capital good all at the same time. The Grossman model, developed by Michael Grossman in 1972, provides such a framework by treating health as something that people decide in part for themselves, rather than something that happens to them. The model provides a powerful set of explanations for a variety of health phenomena, including the link between socioeconomic status and health (Grossman 1972).

3.1 A day in the life of the Grossman model

The Grossman model ties together the health decisions that people make on a day-to-day basis in a framework that encompasses their entire lifespan. We start with the single-period utility function because the function for lifetime utility is built upon that.

Single-period utility

The Grossman model starts with a simplification: in any given period, an individual's utility is based on her health and the other non-health goods she consumes. So the first role health plays in the model is as a *consumption good*. Like the number of chocolate-chip cookies eaten or hours of video games played, one's health contributes directly to utility. An individual's utility for the period t is given by

$$U_t = U(H_t, Z_t) \tag{3.1}$$

where:

- H_t is the level of health, and
- Z_t is a composite good that represents everything else – video games, opera tickets, paintballing, company of friends – that a well-adjusted utility function includes. We refer to Z as the **home good**.

Note that health *care* does not appear explicitly in this utility function, so in this model the number of vaccines received affects utility only through health H (rather than affecting it directly).

While H and Z are distinct contributors to the utility function, there may be occasions when choices made by the individual simultaneously change H and Z. These choices may pose interesting tradeoffs for the individual. For instance, eating a double-double cheeseburger may contribute positively to Z but also clog the individual's arteries and cause a reduction in H. On the other hand, exercising may increase both the home good Z and health H.

Time constraints within a single period

In addition to the possible tradeoffs between H and Z, there are other constraints that limit the individual's ability to gain utility. Perhaps the most important of these is the time constraint – there are only 24 hours in the day. This leads us to the next important piece of the Grossman model: our individual divides her time between exactly four different activities. She can spend her time working, playing, improving her health, or lying in bed sick. In any given period t, the individual has Θ units of time at her disposal, and faces the following time constraint:

$$\Theta = T^W + T^Z + T^H + T^S \tag{3.2}$$

where:

- T^W is time spent working,
- T^Z is time spent playing,
- T^H is time spent improving health, and
- T^S is time spent sick.

In this formula, we suppress the t subscripts on each term to keep the notation simple, but please remember that there is a new stock of Θ units of time to spend in each period, and hence a new constraint. We will include t subscripts when needed for clarity.

Each of these activities plays a different role in the Grossman model and contributes to the individual's utility in a different way. Each hour spent working (T^W) produces income, which can then be used to buy medical care (which contributes to H) or jigsaw puzzles (which contributes to the home good Z). But it is not enough for the individual to simply own a pile of jigsaw puzzles that she never solves. To produce Z, she must actually open the box and piece the puzzle together; that is, she must spend time at play (T^Z). Similarly, she might buy a yoga mat or treadmill with her earned income, but she must spend time using them (T^H) in order to actually produce H.

The time spent sick, T^S, is a different kind of activity. It does not contribute to H or Z, and hence does not increase utility. It *does* impose an opportunity cost, because each hour spent sick is an hour not spent at the gym, or at work, or at the opera. Time spent sick is therefore lost time, and there are only Θ hours in a day. Why would she choose to spend any time sick? She may not have a choice. In the Grossman model, T^S is entirely determined by H and is not a voluntary activity.

Consider *The Simpson's* character Homer, who works at the Springfield Nuclear Power Plant. On Monday, he goes to work (T^W), earning enough income to purchase the latest Troy McClure DVD and a box of day-old donuts. On Tuesday, he decides to skip work, forgoing income in order to spend quality time in front of the television with his donuts and new DVD (T^Z). His activities on Tuesday increase the home good Z, because Homer derives enjoyment from Troy McClure's acting and the taste of jelly donuts. Unfortunately, his day is sedentary – devoid entirely of exercise (T^H) or any physical movement whatsoever. As a result, his health H deteriorates rapidly.

Donuts, a major source of Z *(and negative* H*) for Homer Simpson. Credit: © ksena32 – Fotolia.com.*

By Wednesday morning, Homer is feeling very sick, so he visits his doctor Dr Nick, taking an hour out of his morning to do so (T^H). Dr Nick writes him a note excusing him from work for another day – his advice for Homer is to take placebo medication. The visit and the medication, which cost \$50, improve Homer's H somewhat, but it cannot salvage his Wednesday. Homer is so miserable he cannot even drag himself out of bed to watch Troy McClure for a fifteenth time or eat any of the remaining stale donuts (T^S). For Homer, this is just wasted time. In just three days, Homer has illustrated most of the dynamics of the Grossman model.

Table 3.1 lists examples of activities that fit into each category, and summarizes the purpose and impact of each on the individual's utility.

Production of *H* and *Z*

We focus next on the process of producing H and Z. In the usual economic model of consumer behavior, all inputs into utility are purchased directly on the market. In this case, however, neither H nor Z can be purchased in a store. Instead, the individual must combine market commodities that she purchases with personal time to produce her two inputs into utility. In other words, for both health H and the home good Z, there are two distinct categories of inputs: market goods and personal time.

Table 3.1. *Activities in the Grossman model.*

Activity	Example	Purpose
Working (T^W)	Working at a power plant; playing professional sports; teaching health economics	Earn income to purchase items that will enhance H and Z
Playing (T^Z)	Doing a jigsaw puzzle; going to the opera; logging onto Facebook	Enhance Z
Improving health (T^H)	Jogging; undergoing surgery; beauty rest	Enhance H
Being sick (T^S)	Spending the day home in bed, doing nothing	None; T^S is always wasted time

We have already assigned variables to the personal time inputs that create H and Z: T^H and T^Z respectively. We want to refer to the market goods with shorthand notation as well. Let M be the market inputs for health (for example, health care or exercise equipment), and let J be the market inputs for the home good Z (for example, jigsaw puzzles or video games).

Despite the similarities between H and Z, there is at least one important difference. While Z is a *flow* that is created and consumed each period, H is a *stock* that accumulates or deteriorates from period to period. Decisions made about the individual's health ten years ago affect her today, just as her decisions today will affect her ten years from now. In a sense, the level of H reflects the complete history of past inputs and decisions pertaining to health. By contrast, enjoyment from today's home good is forgotten by tomorrow and does not contribute to tomorrow's Z.

In any given period t, before the individual has allocated any time or money, Z_t starts out at 0. H_t starts out at H_{t-1}, but is modified by the health decisions and purchases the individual makes during period t. The individual treats H_{t-1} as a given during period t, since she cannot go back into the past and change her history of health decisions, which determined H_{t-1}.

Hence, H_t and Z_t are determined by the following production functions:

$$\begin{aligned} H_t &= H(H_{t-1}, T^H_t, M_t) \\ Z_t &= Z(T^Z_t, J_t) \end{aligned} \tag{3.3}$$

where:

- M_t represents market inputs like vaccines and treadmills into health H during period t, and
- J_t represents market inputs like video games and opera tickets into the home good Z.

The market budget constraint

In addition to the time constraint, the individual faces a more traditional budget constraint – she cannot spend more than she earns. Suppose that when the individual works, she earns a wage of w dollars per unit of time. Recall that her total time working in period t is T^W_t, so her total income Y_t for the period t is

$$Y_t = w \cdot T^W_t$$

The Grossman model does not specify how wages w are determined, but presumably the individual's education and other factors determine the wage she faces.

In any given period, the individual can spend her income on two items: market inputs into health (M_t) and the home good (J_t). Let p_M and p_J represent the prices of these two goods. She thus faces the following budget constraint:

$$p_M \cdot M_t + p_J \cdot J_t \leq w \cdot T_t^W = Y_t \tag{3.4}$$

To simplify matters, we assume that the individual cannot save leftover income to spend in future periods. With this additional assumption, the budget inequality becomes an equality:

$$p_M \cdot M_t + p_J \cdot J_t = w \cdot T_t^W = Y_t$$

The time and budget constraints are not independent. They are linked through the individual's decisions about time spent working (T^W). For instance, if the individual is so sick that she has no time to work, she will not be able to earn income to purchase any M or J, the inputs into the production of H and Z.

Sick time and productive time

We turn next to the relationship between health levels H and sick time T^S. The link is intuitive: the healthier the individual is, the less time she spends sick in a given period and the more productive time T^P she has available. Productive time is simply the complement of T^S and is the sum of time spent on the other three useful activities of working, playing, and improving health:

$$T^P \equiv \Theta - T^S = T^W + T^Z + T^H \tag{3.5}$$

In the Grossman model, there are diminishing marginal returns to productive time from health. If a person is already healthy enough to have very little sick time, then additional improvements in health yield little additional productive time. Conversely, if a person is very unhealthy, even small improvements in health can yield substantial decreases in sick time. This is the second role of health in the Grossman model. In addition to its role as a consumption good, health is an *input to the production* of productive time. Figure 3.1 is the illness-avoidance function. It plots the relationship between T^S and H in the Grossman model. T^S falls as health H improves, but the effect of better health on T^S also shrinks as health improves.

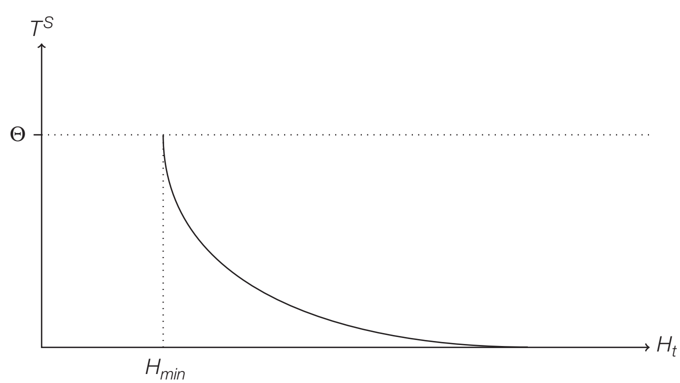


Figure 3.1. ***Illness-avoidance function.*** *As health (H) increases, sick time (T^S) falls. At H_{min}, there is no productive time for any other activities left since health is so low.*

A careful examination of this figure shows that there is a point, labeled H_{min} on the H-axis, such that the individual's T^S equals Θ. This means she is sick for the entire period, with no time left to work, play, or even seek medical care. In this way, the model provides an economic definition of death. The individual has no productive time left, and cannot generate any more health. She remains at H_{min} for all remaining periods and is effectively dead.

Let us go back to the case where the individual is still alive, with $H > H_{min}$. The only way to reduce sick time T^S is to improve health. So any market inputs or personal time dedicated to improving health create extra productive time. This new-found productive time can be reinvested in health-improving activities (T^H), but it can also be put to use as time spent working (T^W) or playing (T^Z). Ultimately, the purpose of reducing sick time is to have more productive time for producing more H and Z.

We have already discussed *two* important roles of health H in any given period. First, it enters directly into utility, as can be seen in equation (3.1). The individual cares about health for its own sake because it feels good to be healthy. Second, it expands the total amount of productive time available to the individual, allowing her to spend more time on things she actually cares about and less time sick in bed. In other words, you need to be healthy to have fun.

There is a third role that health plays in improving utility, but this additional role is not realized within a single period. Because health is a form of *stock*, high levels of health in one period lead to high levels of health in subsequent periods (see equation (3.3)). This third role is sometimes called the investment aspect of health. We will discuss this further in Section 3.3.

Three roles of health in the Grossman model

1 Health is a *consumption good*. It contributes directly to the individual's utility function each period. Being healthy is valuable in and of itself.
2 Health is an *input into production*. It generates productive time T^P which is useful for producing more H and Z.
3 Health is a form of *capital*. Unlike the home good, it endures from period to period. It can accumulate (or depreciate) over time, so improvements in health today can lead to better health tomorrow.

3.2 An optimal day

In the previous section, we discussed some of the key assumptions underlying the behavior of an individual in the Grossman model during a single period. There are many moving parts to the model. The subject of this section is how those parts move together, and how they jointly determine the optimal values of H and Z each period. In the previous section, we maintained a focus on the decisions an individual makes during a single period. Of course, decisions about health in this period have consequences in future periods, because H is a stock. As we will see in the next section, the individual makes decisions with this aspect of health in mind. The optimal levels of H and Z in any given period depend on decisions that are right for a *lifetime*, not just for a single period.

In this section, however, we first suppose that the individual is optimizing utility over just one period in order to demonstrate the major tradeoffs within the model. While this is a simplification, it is one that builds intuition for what follows. In subsequent sections, we relax this simplification and consider optimization over the whole lifespan.

The production possibility frontier for H and Z

A production possibility set traces out all of the possible combinations of H and Z that are attainable given an individual's budget and time constraints. The edge or frontier of this set is called the *production possibility frontier* (PPF) and should be familiar to students of economics. It is helpful to think about what the PPF looks like in the Grossman model, as the constraints facing the individual are not the same as those consumers face in a typical decision model.

In a standard model of consumer decision-making, if a consumer decides to devote all his resources to one good, he can attain a maximal level of that good. For instance, in a model of how consumers split their money between apples and bananas, every apple that the consumer buys results in less money left over to buy bananas. This results in a PPF that resembles Figure 3.2(a). An individual who devotes all resources to H will have no resources left to buy Z. Conversely, an individual who devotes all his resources to Z will have no H. The problem with this typical PPF in the context of the Grossman model is that, as we have seen, an individual with low H will have few resources to produce any Z at all. In the figure, points such as the one labeled "P" are not attainable and therefore should not be within the boundary of the PPF.

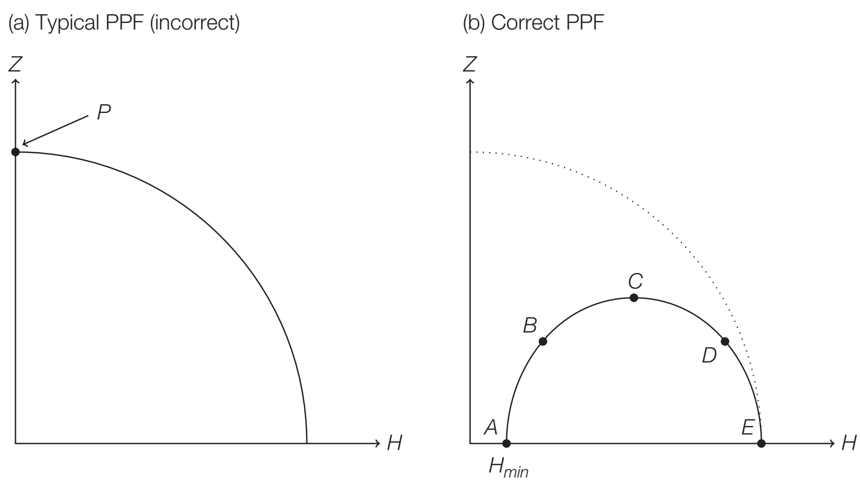


Figure 3.2. *Production possibility frontier (PPF) in the Grossman model. The incorrect PPF (a) describes health and the home good as substitutes. The maximum level of the home good is attained when health is* 0. *The problem with this PPF is that an individual with low* H *will have few resources to produce any* Z. *Points such as the one labeled "P" are not attainable and therefore are not within the boundary of the true PPF (b).*

Figure 3.2(b) presents the correct PPF that is consistent with the budget and time constraints of the Grossman model. It is easiest to see why this frontier is shaped the way it is by examining five extreme points labeled in the figure.

- At point A, the individual is at H_{min}, and hence has no productive time available to work, play, or improve health. As a consequence, he cannot afford any of the home good Z.

- At point *B*, the individual is healthier, and has some time available for productive activities. Since his health is still low, he is on the steep portion of the illness-avoidance function (Figure 3.1). Even small improvements in health yield large increases in productive time. We could call this the *free-lunch zone*: an hour spent increasing health yields more than an hour reduction in sick time. The individual can increase Z without giving up H.
- At point *C*, the free lunch is over. One extra hour spent on health yields exactly one extra hour of productive time T^p. The individual is still not as healthy as he could be, because he is not spending all his time jogging or all of his money on medical care. At point *C*, the individual enjoys the maximum amount of Z possible. If he tries to increase Z by shifting resources from health to the home good, the increase in sick time will outweigh the gain in the resources available for Z production. If he tries to increase Z by increasing H in an attempt to gain more productive time, he will again fail. Increases in health will not produce enough extra productive time to offset the time he must dedicate to improving H.
- Point *D* is in the *tradeoff zone*, which consists of all the points between *C* and *E*. Because the individual is on the flat part of the illness-avoidance function (Figure 3.1), increases in H yield only small decreases in sick time. In order to finance any increase in H, he must shift resources away from Z.
- At point *E*, the individual spends all of his time and money on health, totally ignoring the home good. As a result, Z is 0 and H is at its maximum attainable value.

Picking the optimal H and Z within a period

How does the individual choose the optimal mix of H and Z? Just as in any consumer demand model, the individual picks H and Z to maximize his utility subject to the constraints he faces. Suppose the individual picks a point, such as *A* or *B* in the free-lunch zone of Figure 3.3 (that is, any point on the PPF between *A* and *C*). In that zone, he can simultaneously increase both H and Z simply by shifting resources around. Because of this and the fact that H and Z are positive inputs into his utility function, he never finds it optimal to remain in the free-lunch zone.

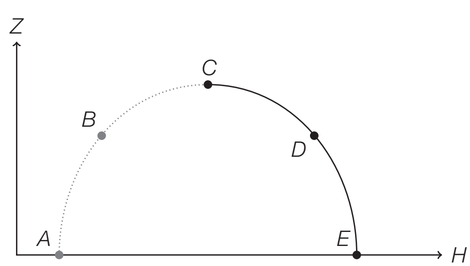


Figure 3.3. *Points* A *and* B *are in the free-lunch zone of this individual's PPF curve, since* C *is strictly better than both. The individual would never choose a point left of* C.

This may be counter-intuitive; why would it be non-optimal to have a free lunch? The answer is that allocations in the free-lunch zone do not take advantage of all opportunities and leave free H and Z on the table. For this reason, we gray out the zone in Figure 3.4 to indicate that no one who values H and Z would ever choose allocations there.

We know that the optimal allocation will be in the tradeoff zone on the PPF; that is, any point between *C* and *E*. The exact allocation he picks depends on his tastes for health and the home good. His utility function (equation (3.1)) describes his preferences over H

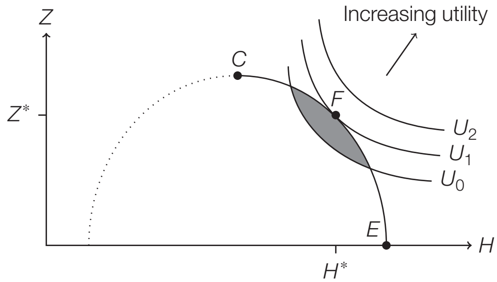


Figure 3.4. *Single-period indifference curves.*

and Z. Recall that in this section we have adopted the simplification that utility is maximized over one period. In the full version of the Grossman model, individuals consider tradeoffs across periods as well. We consider this in Section 3.3.

Recall that an indifference curve maps out the set of allocations that yield the same given level of utility. In Figure 3.4, we plot three example indifference curves. U_2 represents the highest level of utility of these three. The individual would love to be at U_2, but none of the points on that curve are feasible because the indifference curve does not intersect the PPF.

"There is nothing worse than fussiness about one's health, in excess of normal physical training" – Plato, The Republic. *Plato avoided point* E *in Figure 3.4, preferring to concentrate on* Z *(in his case, the search for truth).* *Credit: Zu_09 – iStockphoto.com.*

By contrast, the individual can obtain utility level U_0 by picking an allocation inside the PPF that is on the U_0 indifference curve. This is not optimal though, because any point in the shaded region between U_0 and the PPF would increase his utility. All of these points lie inside the PPF, so they are all feasible.

The allocation represented by F is optimal because it generates the highest utility level possible. Unlike with U_0, there is no room between U_1 and the PPF to improve the allocation. And unlike with U_2, U_1 is actually attainable because it intersects the PPF. In fact, at F, U_1 and the PPF are tangent. H^* and Z^* denote the optimal levels of health and the home good.

For this typically shaped utility function, the individual picks a level of H^* that is less than the maximum. He is willing to give up some health in order to gain utility from other goods. Rather than spend all his time jogging and all his money on checkups, he buys video games and jigsaw puzzles to play with in his leisure time. At the same time, he does not pick the maximum level of Z because he also values better health. This is a key prediction of the Grossman model – that in maximizing their utility, people make tradeoffs that lead to less than maximal health.

This is true for the typical utility function, but the model also allows for exotic preferences. For instance, consider someone who cares only about her health and nothing else. That individual will have indifference curves that are vertical lines, because a change in Z does not affect her utility. Figure 3.5 shows that the optimal allocation for this individual falls at point E, with health at its maximum and Z at 0.

Another possibility is an individual who only cares about Z. This individual will have horizontal indifference curves, since H does not enter his utility function. Figure 3.5 shows that the optimal allocation for this individual will fall at point C, with Z at its maximum. The striking outcome here is that even though H does not enter the individual's utility function, his optimal level of H is still positive. Recall that there are two possible motives for choosing H in the single-period model: because it directly enters the utility function and because it increases productive time. In this case, the individual picks a positive level of H solely because of the latter motive.

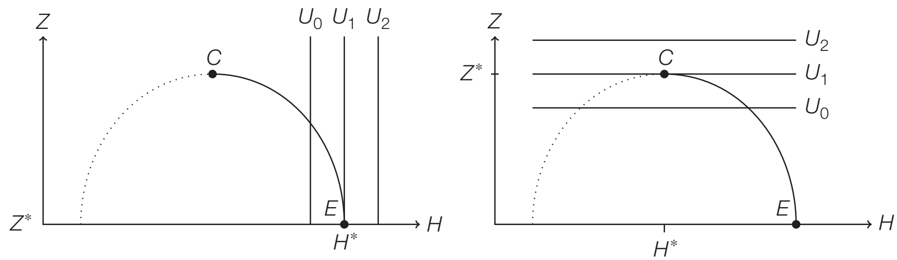


Figure 3.5. *Exotic indifference curves.*

The labor–leisure–health improvement tradeoff

The individual has to navigate a complicated set of tradeoffs in deciding how to spend his time. In this subsection, we take a closer look at these tradeoffs and how the individual's utility function and constraints interact. In a sense, we already looked at this tradeoff when we considered the tradeoff between H and Z. In order to allocate resources between H and Z, the individual first allocates his time across the four activities: work (T^W), play (T^Z), health improvement (T^H), and illness (T^S).

This is a complicated decision process in any given period, but the problem is simplified by the fact that the individual's prior health, over which he has no current control, plays a key role in determining current health. Current decisions about health will affect his future health, but that is little help to him in the current period.

The total productive time he has available, T^P, is thus determined by his health. Figure 3.6 shows this relationship. This figure is derived directly from Figure 3.1, where we plotted T^S against H. Remember that $T^S + T^P = \Theta$, so that both figures are plotting the same relationship between health and productivity. Total productive time, not surprisingly, is increasing in health, but there are diminishing marginal returns.

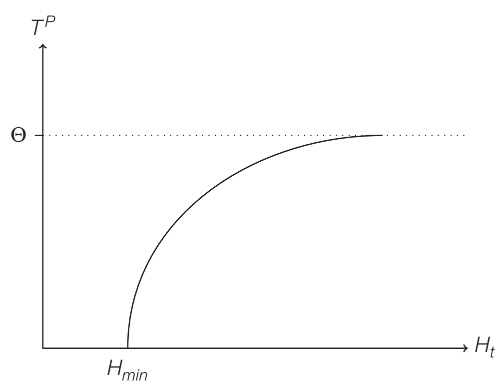


Figure 3.6. *At low levels of health, small improvements reap large increases in productive time,* T^P, *and large drops in sick time,* T^S. *But the same amount of health improvement at high levels of health returns a much smaller gain in* T^P. *Thus, marginal returns to health are diminishing.*

So we can simplify the individual's choice. He now has T^P available to him to allocate among three activities: work, play, and health improvement. This is still a complicated choice, but easier than the choice he previously faced.

Figure 3.7 shows the set of possible allocations for his productive time, plotted in three-dimensional space. At point A, the individual is a workaholic: he devotes all of his productive time to work and none to play or health improvement. Similarly, points B and C represent the extreme healthaholic and playaholic, respectively.

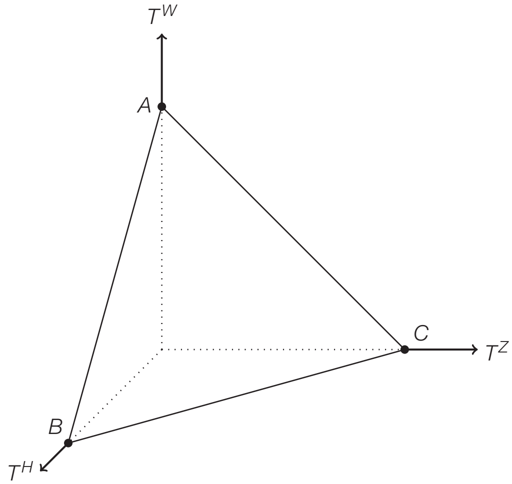


Figure 3.7. *After suffering time sick, the individual allocates productive time* $T^P = \Omega - T^S$ *between labor* T^W*, play* T^Z*, and health improvement* T^H*. The decision is shown here as a three-dimensional tradeoff. Later, we hold* T^H *fixed and study the decision in two dimensions.*

Visually, it is easier to think about decisions in two dimensions. So let us suppose the individual has already made a final decision about time spent on health improvement T^H. We are doing this to make the tradeoffs easier to see, but the reader should always remember that T^H, T^W, T^Z, and all the other decision variables are chosen simultaneously in the Grossman model.

If we assume that the individual has already decided T^H, then the individual has only two ways to spend his remaining time: labor T^W and leisure T^Z. Figure 3.8 shows this tradeoff, along with indifference curves that illustrate his preferences between work and play. He must balance his desire to play, which helps produce Z, with the necessity to work, which allows him to buy inputs into H and Z. The shape of the indifference curve U_0 in this figure is therefore indirectly derived from his utility function for H and Z. This indifference curve is the "same" indifference curve we saw in Figure 3.4, projected onto a different set of axes.

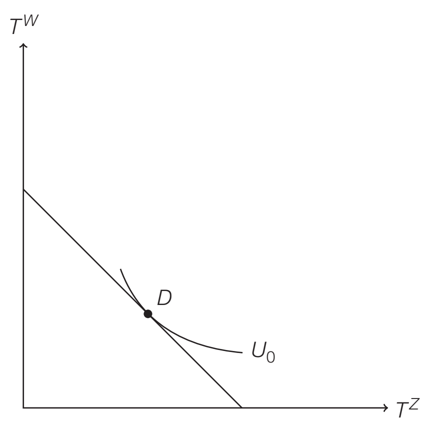


Figure 3.8. *Once sick time* T^S *and health-improvement time* T^H *are set, the individual allocates the remaining time between work* T^W *and play* T^Z*.* D *is the optimal allocation of time, where the indifference curve* U_0 *is tangent to the time constraint.*

The individual chooses the optimal levels of work and play using a familiar mechanism. He picks T^W and T^Z at the point where his indifference curve is tangent to the constraint he faces. In Figure 3.8, this is shown at point D. This individual in question is neither a workaholic nor a playaholic because he allocates some time to each activity.

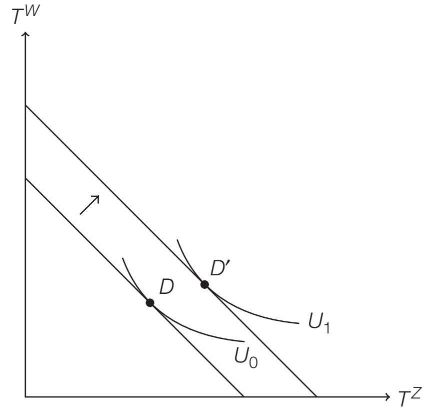


Figure 3.9. *When health improves, more productive time is available. The individual can spend more time both at work and at play. Graphically, the individual's time constraint shifts outward, and allows him to reach* D′, *a new optimum with more* T^W, T^Z, *and utility than* D.

This analysis helps us see more clearly the value of health improvement. Imagine that the individual had spent previous periods jogging and quitting smoking instead of his usual slothful routine of video games and jelly donuts. He would have entered this period in better health, and consequently, would have suffered lower T^S and had more productive time at his disposal. Figure 3.9 shows the payoff of this prior health investment. The constraint shifts outward, allowing more time for both labor and leisure.

Improved health from previous periods and reducing T^S is not the only way to gain more time for labor and leisure. Alternatively, the individual could sacrifice some time spent on health improvement T^H in this period in order to free up time for work and play. Of course, this will have repercussions in future periods on his health and other outcomes.

3.3 Extending Grossman from cradle to grave

So far, our examination of the Grossman model focused on the life of an individual in a single period. This allowed for an introduction to the basic mechanics of the model, and demonstrated the important tradeoffs (between health and other goods, between useless sick time and costly health investment, between labor and leisure). But as we also hinted, health is a stock; decisions made in the past impact today, and decisions today have implications in future periods. In this section, we examine the full multi-period version of the Grossman model and study how the individual navigates the tradeoffs between health, work, and play over an entire lifespan.

The multi-period utility function

We are finally ready to present the full range of our individual's preferences. He values health and home goods in every single period of his life:

$$U = U(H_0, Z_0, H_1, Z_1, \ldots, H_{\Omega-1}, Z_{\Omega-1}, H_\Omega, Z_\Omega)$$

where:

- H_t is the level of health in period $t = 0 \ldots \Omega$,
- Z_t is the amount of the home good in period t, and
- Ω is the length of the lifespan in periods.

As we shall see later, Ω is actually chosen by the individual.

We consider a version of this utility function in which the individual separates decisions in each period. With this functional form for utility, each period produces a *flow* of utility $U(H_t, Z_t)$ that contributes to the overall lifetime utility:

$$\begin{aligned} U &= U(H_0, Z_0) + \delta U(H_1, Z_1) + \delta^2 U(H_2, Z_2) + \cdots + \delta^{\Omega} U(H_{\Omega}, Z_{\Omega}) \\ &= \sum_{t=0}^{\Omega} \delta^t U(H_t, Z_t) \end{aligned} \tag{3.6}$$

where $\delta \in (0, 1)$ is the individual's discount factor. Since δ is between 0 and 1, when δ is raised to a power it becomes smaller. We need this discount factor to represent the fact that the individual values the current utility flow more than he values the same amount of utility flow in any future period.

Health as an investment good

Now that we have introduced the multi-period utility function, we are ready to discuss the third role of health (recall that the first two roles of health were as a consumption good and as an input in the production of healthy time). Health is a form of human capital akin to knowledge or education. All capital goods have a few things in common. They store value from investments in previous periods, but they also depreciate in value over time.

Health is no exception. The human body, like a car or pizza oven, may last for a long time but suffers the typical wear and tear that comes even with careful use. Let γ be the rate of depreciation; this measures how fast health H dissipates from period to period.

We are now ready to revise the production function for health (see equation (3.3)) so that it reflects the depreciation of health over time:

$$H_t = H\left((1-\gamma)H_{t-1}, T_t^H, M_t\right) \tag{3.7}$$

As we will see, depreciation of health plays a critical role in determining the optimal level of health investment.

Return to health capital

Health is a capital good, which is another way of saying it is an investment with its own rate of return. However, health has a particular feature – at low levels of health, small investments have enormous returns to productive time.

Let us consider a thought experiment: What happens to an individual's lifetime utility when her health in any one period is magically increased by a small amount? This increase in health has repercussions on her health and utility level in each period for the rest of her lifespan. The total increase in lifetime utility, which is the "return" to this increase in health, depends on her starting level of H. The marginal lifetime returns to health are high at low levels of health and low at high levels of health because of the diminishing marginal returns to health (recall Figure 3.6).

Figure 3.10 shows how the marginal lifetime returns to health change with the initial level of H. This curve is called the **marginal efficiency of capital (*MEC*)** because it indicates how efficient each unit of health capital is in increasing lifetime utility. Recall that these returns are measured over the lifespan and include all benefits of health. Ironically, the highest returns to health are available when the individual is dying, $H \approx H_{min}$. Even then, an investment to health does not have infinite returns. The *cost* of making that investment may be so high that even high returns do not justify investing.

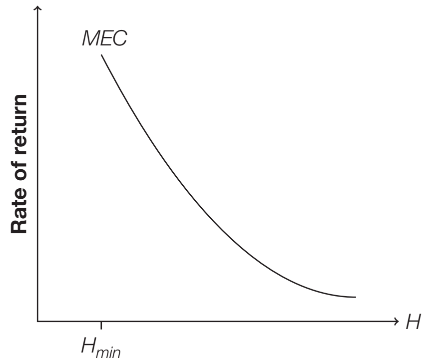


Figure 3.10. *The marginal efficiency of health capital curve (*MEC*) captures all lifetime return from a marginal investment in health at any given level of health stock. That the* MEC *curve is downward-sloping reflects the diminishing marginal returns to health.*

What are the costs of investing in H? First, there is the opportunity cost; the individual forgoes putting her resources in other market investment options. Let us suppose these alternative market investments pay an interest rate r. For ease of discussion, we earlier assumed the individual cannot save income from period to period, but here we relax that assumption to make clear that there is an opportunity cost to using health as a savings vehicle.

Depreciation due to aging γ acts as a second type of cost of investing in H. Suppose the individual invests in health. In order to guarantee the same rate of return as the market investment opportunity, health must pay a return of at least $r + \gamma$. If the return were any less, then depreciation lowers the effective return to health below r, making the market opportunity more attractive. Thus, the market rate of return to the alternative investment plus depreciation $(r + \gamma)$ is the effective price of health capital.

Most students of economics are used to thinking about opportunity costs, but it may be confusing to think of depreciation as a *cost* of capital as well. In what sense does health get more expensive when γ rises? Think about the fact that, in order to have the same amount

Actually, the Grossman model is named after economist Michael Grossman, who developed it as part of his PhD dissertation.

Credit: Allen Cox.

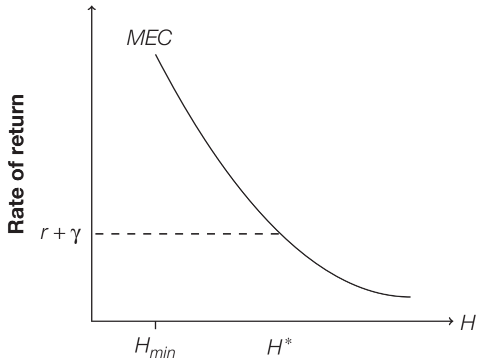


Figure 3.11. *Optimal investment in health depends both on the opportunity cost of forgone market investment opportunities* r *and the individual's depreciation of health* γ. *High* r *and* γ *lower the individual's optimal level of health.*

of health later, you need to buy more now if the depreciation rate is high. The depreciation rate acts like a kind of continual tax on any health the individual produces.

The *MEC* curve determines the optimal amount of health H^* for the individual. On a traditional demand curve for a given good, the quantity associated with a particular price is the optimal demand for the good at that price. Similarly, the *MEC* curve shows the optimal health level associated with the market price of health investment, $r + \gamma$. At this price level, the individual optimally chooses H^*. And at H^*, the marginal cost of health investment $(r + \gamma)$ balances the marginal benefit of health investment.

3.4 Comparative statics

The measure of a model is its ability to represent complex realities and make predictions. The Grossman model is best known for providing economic explanations for two well-known empirical health phenomena: better health among the educated, and declining health among the aging. We use comparative statics to study the health of the college graduate versus that of the high-school dropout, and the health of the twentysomething versus that of the senior citizen. This side-by-side analysis allows us to better understand the power of the Grossman model.

Education and the efficiency of producing health

There is much evidence of a relationship between health and socioeconomic status (SES). People who are better educated or wealthier tend also to enjoy longer life expectancies and fewer health problems. This correlation is known as the **SES health gradient** and will be studied further in the next chapter.

There are several different explanations for the gradient advanced by sociologists, doctors, and economists. The Grossman model posits one: the gradient arises because the well-educated are more efficient producers of health. For any given hour dedicated to T^H or dollar devoted to purchasing health-related goods M, a college graduate reaps more health improvement than a high-school dropout. This difference may take the form of a better understanding of a doctor's instructions or more sophistication about purchasing medicine. In the language of the Grossman model, the more-educated person gains more health stock for each unit of health investment.

Graphically, this increased efficiency manifests as an upward shift of the marginal efficiency of health capital curve *MEC*. Because better-educated people are more efficient health producers, they receive higher returns from health investment at any H. In Exercise 13, the student will have the opportunity to further explore the effect of changes in wage on health.

Figure 3.12 displays the MEC curve of two different individuals, one a college graduate (MEC_C) and the other a high-school dropout (MEC_H). The two have different education backgrounds but are otherwise exactly the same: they share the same preferences for labor and leisure, the same taste for Z and H, and the same depreciation rate γ and return for market investments r. Hence, if not for their education differences, the Grossman model would predict their lives would be identical.

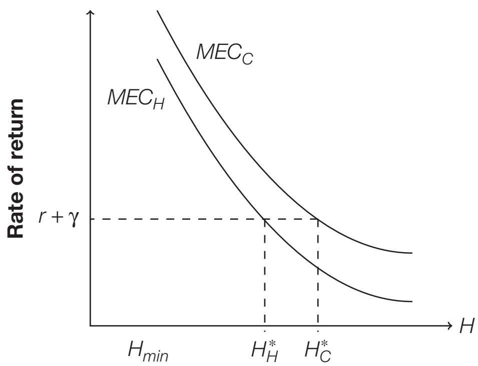


Figure 3.12. *Empirical evidence consistently finds that the more-educated enjoy better health than the less-educated. One hypothesis for explaining this difference is that the college-educated are more efficient producers of health. This is graphically depicted as a higher* MEC *curve for the college graduate (*MEC_C*) compared with the high-school dropout (*MEC_H*).*

But their education differences matter a lot. In Figure 3.12, the college graduate's higher MEC_C means that her optimal level of health H_C^* is higher than the high-school dropout's H_H^*, even though both face the same cost of capital $r + \gamma$. Because the college graduate is more efficient in producing health, it makes sense for her to invest more in her health. The high-school dropout, on the other hand, does not produce health as efficiently, so investments in health offer lower returns. It makes more sense for him to focus on Z.

This disparity in education and resulting differential in health has major implications. Not only will the college-educated person be healthier throughout her life, as empirical evidence suggests, but, as we see later in this section, she also enjoys a lengthier lifespan.

Aging and endogenous death

A second important payoff of the Grossman model is that it provides an economic explanation for why health optimally deteriorates with age. We might ask why not make investments so that H stays high forever? The Grossman model predicts that it is too costly to stay forever young.

There is a basic biological reality of aging: whatever health assets a person has (strong bones, unclogged arteries, plentiful neurons) not only tend to diminish over time, but tend to diminish at a *faster rate* as he ages. In economic terms, the depreciation rate γ does not stay constant over a lifetime. As the individual ages, his γ also increases.

Now consider what happens when the individual's γ begins to rise as he ages. The cost to capital $(r + \gamma)$ rises, and he would have to invest more and more resources to maintain the same level of health over time. As the cost of health rises, the aging individual will be

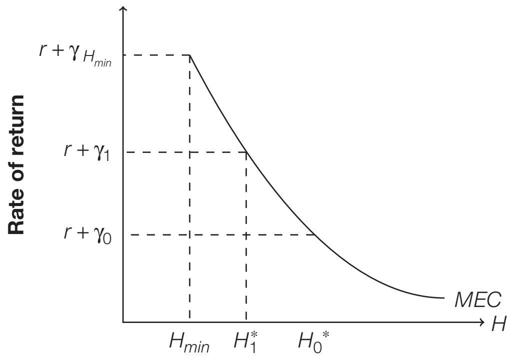


Figure 3.13. *As the individual ages, his depreciation of health increases from γ_0 to γ_1 and finally, to $\gamma_{H_{min}}$. Such a high depreciation level means any investments in health decay immediately, so the individual may not want to invest in health at all. This is the Grossman model's treatment of death.*

less willing to spend his productive time and money on H and will start devoting a greater proportion of his dwindling resources to Z.

Figure 3.13 shows that, as the cost of capital $r + \gamma$ rises with age (since γ is always increasing), the optimal health H^* will decline. As γ continues its inexorable climb, eventually H^* falls to H_{min}.

At H_{min}, all available time is lost to T^S. The individual has no time to spend on labor, leisure, or health improvement, and therefore cannot produce any H or Z ever again. The reasonable question at this juncture is why any rational actor might choose to set $H = H_{min}$ for any period. How could death possibly be utility-maximizing?

Consider the alternative: Under what circumstances would it be optimal to live forever? Imagine, contrary to biological realities, depreciation of health stays at a γ of 5% from age 0 onward. It is possible to show that, in the Grossman model, a constant depreciation rate results in an equilibrium level of health H_e^* that never changes. In each period, 5% of H drains away, but the individual invests exactly enough resources to replenish his supply of H right back up to H_e^*. After this equilibrium is reached, every future period is identical. The individual lives out an infinite lifespan at H_e^*. This is a pretty outlandish result, even for an economics model. Unfortunately, we would need a fountain of youth for this happy result.

The key insight is that increasing depreciation γ makes it less and less attractive to invest scarce resources in health. To see why the high depreciation rate makes health investments less appealing, consider what would happen if a student knew that he would forget 99% of whatever he studied before a big exam. How would that affect his willingness to hit the books? This "why bother?" attitude dissuades aging individuals from investing in health in the Grossman model.

Fans of the critically acclaimed 2006 film *Little Miss Sunshine* may recall the free-wheeling, foul-mouthed Grandpa character with a destructive heroin habit. At one point in the movie he makes a comment that exemplifies the logic of the Grossman model. After his family discovers his heroin addiction, he counsels his grandson against the habit:

GRANDPA: Don't you get ideas. When you're young, you're crazy to do that s***.

FRANK: What about you?

GRANDPA: Me?! I'm old! You get to be my age - you're crazy *not* to do it.

Neither the Grossman model nor the authors of this textbook endorse recreational drug use among the elderly.[1] Nevertheless, the Grandpa's life choices indicate an understanding of health as a depreciating capital commodity. It seems that, rather than making costly investments (e.g. abstaining from heroin) to maintain a health stock that is dissipating rapidly anyway, he opts to shift his resources to producing more Z (e.g. a heroin-induced high) even at the expense of his health.

3.5 Unifying the Grossman model

At this point, we have introduced four different graphs, each of which illustrates a different tradeoff or constraint in the Grossman model.

- The **production possibilities frontier** (PPF) in Figure 3.4 demonstrates the tradeoff between health production and home good production.
- The **health production** schedule in Figure 3.6 relates health to availability of productive time T^P for work, play, and health improvement.
- The **labor–leisure** graph in Figure 3.7 illustrates the tradeoffs in allocating finite time between work and play.
- The **marginal efficiency of capital** (MEC) curve in Figure 3.11 plots the marginal lifetime utility returns against initial levels of health.

The aim of this section is to tie together these four diagrams. The preferences they illustrate are intimately related in a single, multi-dimensional utility function that underlies all four graphs. The indifference curves in the PPF diagram and the work–play tradeoff diagram may seem unrelated, but they actually reflect the same underlying tastes for H and Z. The tradeoffs and constraints that people face are also multi-dimensional.

Figure 3.14 combines these four graphs and aligns the axes so as to illuminate the connections between the components of each figure. All the diagrams in this figure are snapshots of an optimization problem taking place in high-dimensional space. Each graph in Figure 3.14 is a projection of an aspect of this problem onto a two-dimensional plane. Simultaneously solving for the optimum in each diagram results in the bundle of H_t, Z_t, T_t^H, T_t^Z, T_t^W, T_t^S, M_t, J_t that maximizes the individual's lifetime utility.

To better see how the diagrams interrelate, Figure 3.14 shows the effect of an increase in age discussed in Section 3.4. It depicts the decisions of the same individual at two different stages of his life: the "young" period, denoted by subscript Y, and the "old" period, denoted by subscript O.

Recall that the direct result of the aging process is an increase in depreciation rate γ, which is represented in part (a). As a result of the higher depreciation, the optimal level of H decreases as the aging individual shifts investments away from health. Accordingly, the optimal level of health of the individual in the old period H_O^* is less than H_Y^*.

This lower level of health inevitably results in lower productive time, as the individual suffers more sick time in the old period as shown in part (b). The decline in productive time in turn limits the available time for work and play as shown in part (c). Having less productive time also reduces the time and money available to the individual for

1 Indeed, the Grandpa reaches H_{min} before the final scene.

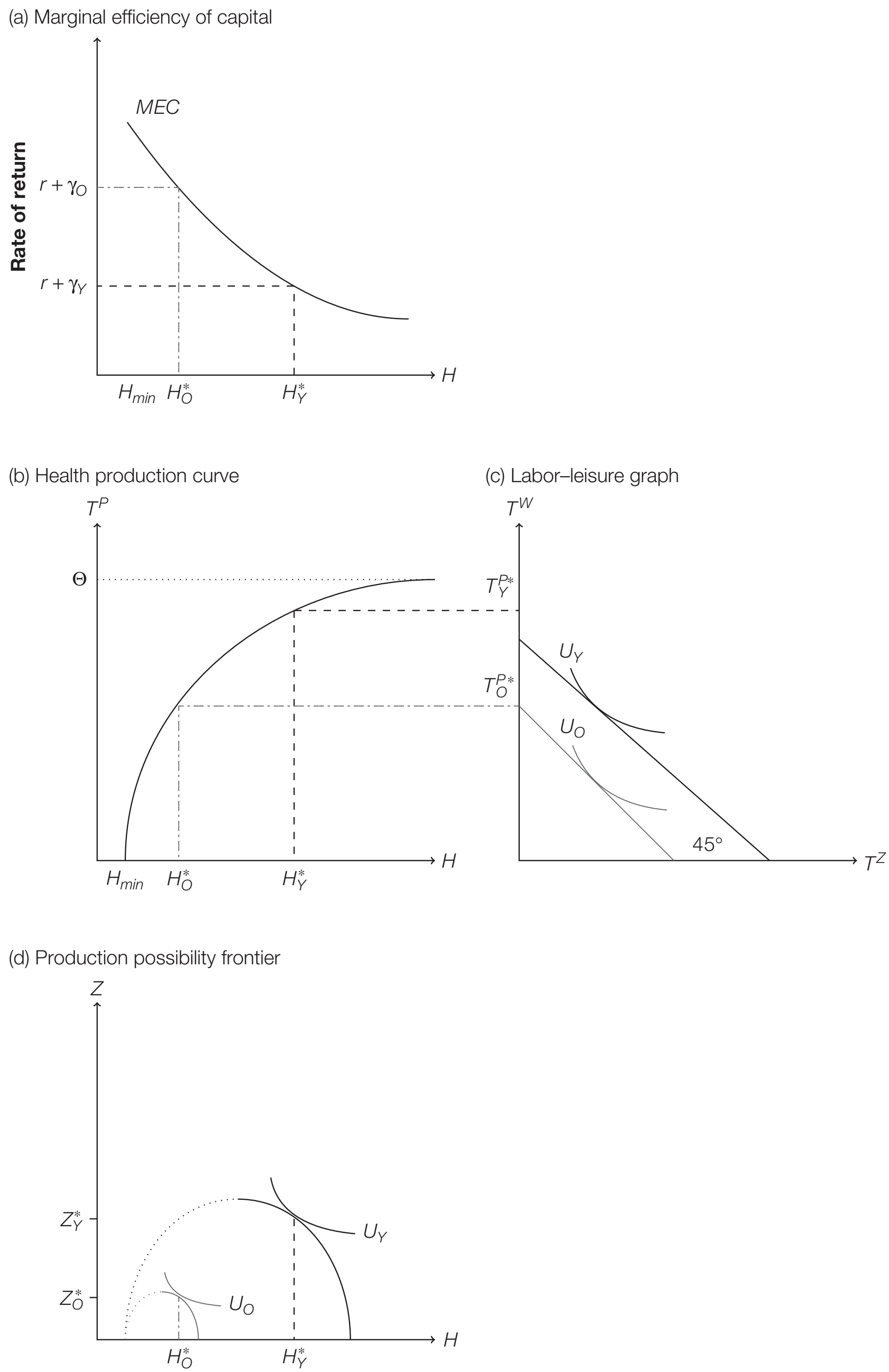


Figure 3.14. *The combined Grossman model shifting with age.*

production of H and Z. This shrinks inward the individual's PPF and results in a lower level of overall utility (shown in part (d)).

3.6 Conclusion

The Grossman model begins with the premise that health is a *choice*, at least when considered over a long period of time. People may start with different endowments, and in any given period may not have a large degree of control, but over time small decisions add up to large changes in health. This idea, combined with declining marginal productivity of health investments and binding time constraints, leads to powerful predictions about health and aging, health and socioeconomic status, and labor–leisure tradeoffs, among others. Our task in future chapters is to examine whether the empirical evidence is consistent with these predictions.

3.7 Exercises

Comprehension questions

Indicate whether the statement is true or false, and justify your answer. Be sure to cite evidence from the chapter and state any additional assumptions you may need. Review the basic assumptions of the Grossman model before answering these questions.

1 In real life, investments in health can generate long-lasting benefits, but the Grossman model neglects this aspect of health.

2 In the framework of the Grossman model, an individual's level of health is completely controlled by her actions. Thus, in any given period, an individual is unconstrained in her choice of health status.

3 In the Grossman model, the marginal efficiency of investment in health care declines as health improves.

4 Aging shifts the marginal efficiency of investment in health curve inward.

5 An hour spent exercising always pays for itself by decreasing the time spent sick by more than an hour.

6 Assume the PPF is as pictured in Figure 3.3. People might choose point E as their optimum even if they value the home good Z.

7 In the Grossman model, optimal health status declines with age.

8 The fact that older people spend more on health care is evidence against the Grossman model, which predicts that spending will decline as δ increases.

9 People who drop out of high school are able to produce more health than college graduates because they have more free time to invest in health production.

10 According to the Grossman model, people choose an optimal time to die (barring any unforeseen accidents).

Analytical problems

11 The Grossman model envisions consumers deciding between investments in health H and investments in home goods Z. Figure 3.15 depicts a typical consumer's production possibility frontier for health and home goods.

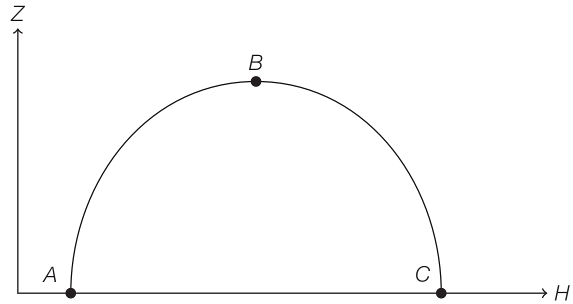


Figure 3.15. *The PPF in the Grossman model.*

- **a** Succinctly describe why the graph is shaped the way it is between points A and B.
- **b** Succinctly describe why the graph is shaped the way it is between points B and C.
- **c** Would any consumer with typical preferences ever pick a point on the graph between A and B? Explain succinctly (using Figure 3.15) why or why not.

12 Suppose a new miracle pill is discovered that increases both the marginal health effects of health investment (at any given level of health investment) and the maximum level of attainable health from H_{max} to a higher H_{max}.

- **a** Draw the old PPF before the discovery of the miracle pill.
- **b** On this same graph, draw a new PPF that corresponds to the description of the miracle pill.
- **c** How will the miracle pill affect H^*?
- **d** How will the miracle pill affect the rate of jogging?

13 **Differences in wage levels.** Suppose individual A received a much better education than individual B, and consequently earns twice as much per hour of labor.

- **a** If both individuals work 40 hours a week, who will have greater H^*? Why?
- **b** If both individuals work enough hours so as to earn exactly $50,000 per year, who will have greater H^*? Why?
- **c** Draw a set of axes labeled T^W and T^Z that resemble Figure 3.8. Draw a line labeled A that shows individual A's time constraint during a given period (say, one month), and draw another time constraint for individual B. Explain why and how they differ due to choices in previous periods.
- **d** On this same set of axes, draw representative indifference curves for both individual A and individual B, including ones that lie tangent to their respective time constraints. Based on the way you drew your curves, does individual A earn less than twice as much or more than twice as much as individual B?
- **e** Explain briefly why your answer to Exercise 13(a) may change if you found out that individual A had to spend ten times as many hours on homework during her schooling years as individual B did.
- **f** Explain why it is optimal for individual B to invest less in health when she is already less healthy than A. Why does she not invest more to "catch up"?

14 True or false? According to the Grossman model, if a new drug were discovered that eliminated the steady deterioration of health that accompanies aging – but does not eliminate sudden events like heart attacks or being hit by a bus – then the demand for jelly donuts, french fries, and physical activity in the presence of buses would decline. Justify your answer.

15 How does aging change the shape or size of the PPF in the Grossman model from period to period? Draw a graph to demonstrate the effect of aging, and include a short paragraph of text justifying the changing shape or size of the PPF.

16 **Nutritional economics**. Suppose we are considering a hungry individual in the Grossman model deciding what to have for dinner. His options are listed in Table 3.2. Each dish has an effect on the level of the home good Z and health H.

Table 3.2. *Meal options in the Grossman model.*

Meal	Home good (Z)	Health (H)
Steak and eggs	$+7$	-2
Kale salad with broccoli	-2	$+5$
Entire box of cookies	$+10$	-20

a Suppose the diner's single-period utility function is as follows:

$$U = 3Z + H$$

If the diner is trying to maximize his single-period utility, and he can only select one item from Table 3.2, which meal would he choose?

b A miracle pill is discovered that halves the negative health impact of cookies. How does this impact the diner's choice?

c What effect does the miracle pill have on the diner's health H? Interpret this result. Does this mean the diner would be better off without the miracle pill?

d If the diner is instead trying to maximize his lifetime utility and not just his single-period utility, how might your answer to Exercise 16(a) change? Is he likely to value Z or H more in the lifetime context than the single-period context? Explain your answer, and be sure to invoke the concept of a capital good.

Essay questions

17 One curious finding from the RAND Health Insurance Experiment was that the rate of treated bone fractures per capita was higher in the group of families that had been assigned to the free insurance plan, compared with those in the high copayment plans. Concisely describe how the Grossman model might explain the fact that people facing higher prices for health care would break bones less often. Be sure to discuss the concept of marginal efficiency of health investment.

18 **Munchausen's syndrome**. Munchausen's syndrome is a psychiatric disease first recognized by doctors in the 1950s. Sufferers will feign unusual medical symptoms and seek out the most complicated treatments and procedures, typically out of a desire to gain the sympathy and attention of family, friends, and medical professionals. In some sense, we could say that health care enters into the utility function of the afflicted. As much as most people viscerally dislike sitting in a doctor's waiting room or undergoing surgery at a hospital, people with Munchausen's often cannot get enough.

Imagine an individual in the Grossman model who suddenly develops Munchausen's syndrome. How would this affect her optimal level of H^*? Explain your answer, and make sure your explanation discusses the three roles of health in the model.

19 A recent paper entitled "Inheritances, health, and death" by Beomsoo Kim and Christopher Ruhm (NBER Working Paper No. 15364, 2009) reports the following:

> We examine how wealth shocks, in the form of inheritances, affect the mortality rates, health status and health behaviors of older adults, using data from eight waves of the Health and Retirement Survey (HRS). Our main finding is that bequests do not have substantial effects on health, although some improvements in quality-of-life are possible. This absence occurs despite increases in out-of-pocket (OOP) spending on health care and in the utilization of medical services, especially discretionary and non-lifesaving types such as dental care. Nor can we find a convincing indication of changes in lifestyles that offset the benefits of increased medical care. Inheritances are associated with higher alcohol consumption, but with no change in smoking or exercise and a possible decrease in obesity.

Interpret these findings using the Grossman model as a framework. In particular, comment on how exogenous income shocks change decisions about health status in the Grossman model. Is the evidence in this report consistent or inconsistent with your interpretation of the predictions of the Grossman model? Does this report support or contradict the theory that one's wealth level determines health status?

Students can find answers to the comprehension questions and lecturers can access an Instructor Manual with guideline answers to the analytical problems and essay questions at **www.palgrave.com/economics/bht**.

4 SOCIOECONOMIC DISPARITIES IN HEALTH

How long will you live? The answer to that question has a lot to do with your standing in society – financial, social, and otherwise. Understanding and explaining the links between socioeconomic status (SES) and health has attracted the efforts of economists, sociologists, epidemiologists, public health researchers, and biologists alike. Few other areas of health economics have garnered as much attention in recent years.

All this scholarly attention to socioeconomic health disparities is merited because they are incredibly pervasive. Health disparities appear in every country on Earth, and in different historical settings as well. We show evidence of health disparities across education, race, employment grade, income, and birth weight. We will even present evidence that health disparities exist in non-human societies.

While the widespread existence of health disparities may be an obvious point, the reasons behind these disparities are a rich source of scholarly investigation and controversy. Researchers have proposed many theories to explain the link between SES and health. Some theories emphasize the effect of SES on health; others emphasize the opposite pathway from health to SES; and yet others emphasize alternate variables that explain the connection between SES and health. Figure 4.1 shows the different causal pathways that might underlie these findings.

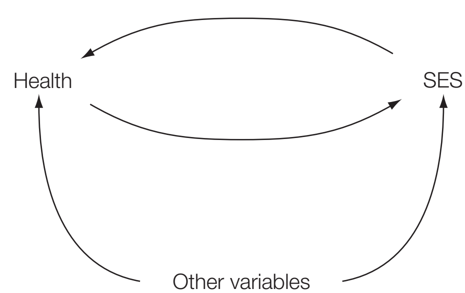


Figure 4.1. *Causal relationship between health, socioeconomic status, and other variables.*

While many theories of health disparities exist, this chapter highlights six prominent theories implicating various factors: adherence to medical advice, early life events, income levels, the stress of being poor, work capacity, and impatience. In Chapter 5, we discuss one other prominent theory: discrimination by doctors. While Figure 4.1 provides one useful way to organize these theories, it is helpful to also consider them through the lens of an economic model. The Grossman model discussed in Chapter 3 provides just such a structure, and we will use it as an organizing principle throughout this chapter.

As with much of health economics, understanding how the evidence relates to economic theory has implications for health policy. For instance, if health disparities arise *in utero* (that is, before people are even born), the optimal policy response will be very different than if health disparities only arise later in life.

4.1 The pervasiveness of health inequality

Health inequality exists in every society. The prevalence of health disparities across different strata is not a phenomenon unique to a particular country, nor is it unique to our time or even to our species. Here, we document the evidence for health disparities in a variety of settings.

Perhaps the most basic measure of health status for a population is survival or life expectancy. Imagine a large population of people, all the same age. As they get older, some of the people in this group will start to die off, while others live on. We can depict the plight of this population with a *survival curve*, a graph that tracks the fraction of the group still alive at each age.

Figure 4.2 shows two such survival curves for American males. One curve (solid line) is calculated for college graduates, while the other curve (dotted line) is calculated for high-school dropouts. The figure shows that 18-year-olds who will eventually graduate from college are likely to survive longer than 18-year-old high-school dropouts. Imagine a high-school class with 100 college-bound young men and 100 other young men who will drop out before graduation. These curves suggest that about 85 of the college graduates would live to attend their fiftieth reunion, but only 60 of the high-school dropouts would survive to that time. Strikingly, the college graduates are 25 percentage points more likely to survive to age 68.

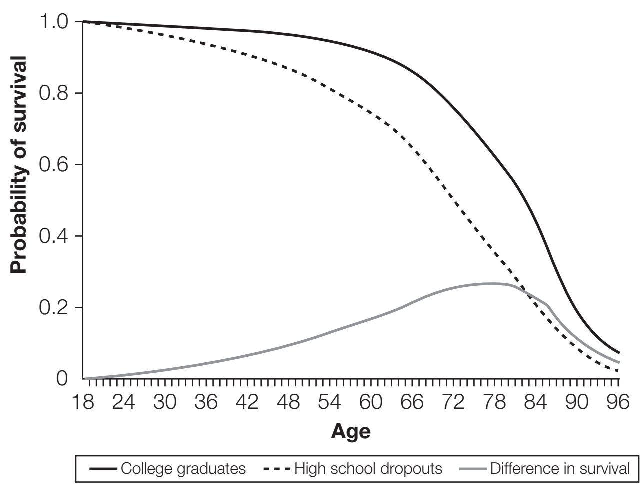


Figure 4.2. *Male survival curves by educational attainment.*

Source: Figure 1 from Bhattacharya and Lakdawalla (2006). Reprinted with permission from Elsevier.

Historical health disparities

Such mortality disparities have existed at least since the beginning of the industrial age and probably much longer. Demographer T. H. Hollingsworth (1965) tracked the life expectancy at birth among members of British ducal families between 1750 and 1900 (Figure 4.3). At the beginning of this period, British nobility had the same average life expectancy as the rest of the British population. By 1900, the children of British dukes could expect to live nearly twenty years longer than their commoner counterparts. Antonovsky (1967) cites similar studies with diverse data sources ranging from 1820s France to civil-war-era Rhode Island, all of which find life expectancy or mortality disparities between the rich and the poor.

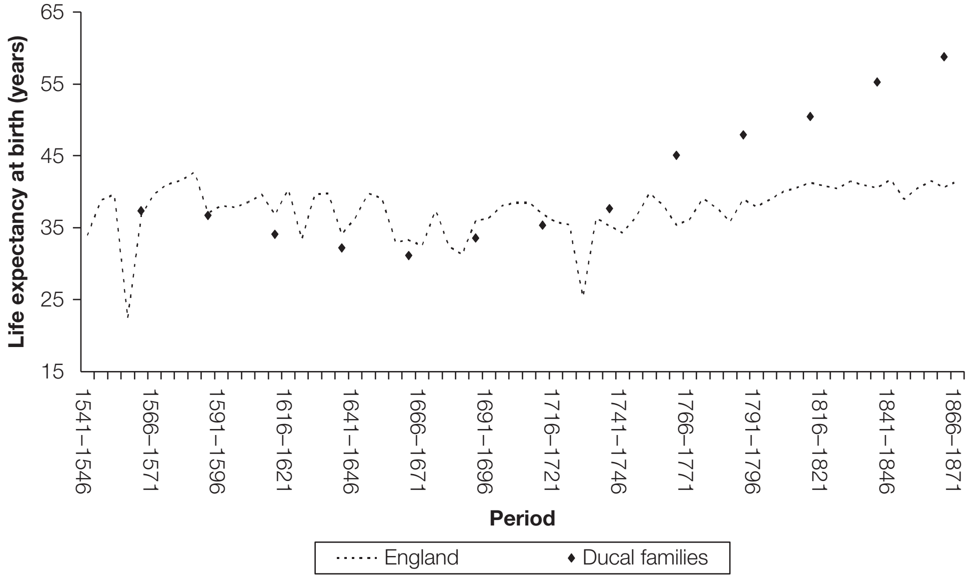


Figure 4.3. *Mortality rate among British ducal families and commoners.*

Source: Reproduced from B. Harris (2004). Public health, nutrition, and the decline of mortality: the McKeown thesis revisited, *Social History of Medicine*, 17(3): 379–407, by permission of Oxford University Press. Original data from T. H. Hollingsworth (1965) and Wrigley et al. (1997).

Disparities across income levels

Mortality or life expectancy is an extreme measure of health; two people living out the same lifespan could enjoy very different levels of health. In fact, health disparities emerge almost no matter how health is measured. For instance, another commonly used measure of health is *self-reported health status*, usually delineated on a scale ranging from 1 (poor health) to 5 (excellent health).

High-income individuals routinely self-report better health status on this scale than low-income individuals. Figure 4.4 shows nationally representative evidence on this point from the United States. Family income is plotted against average self-reported health, and each line represents a different age group. In the graph, higher numbers on the vertical

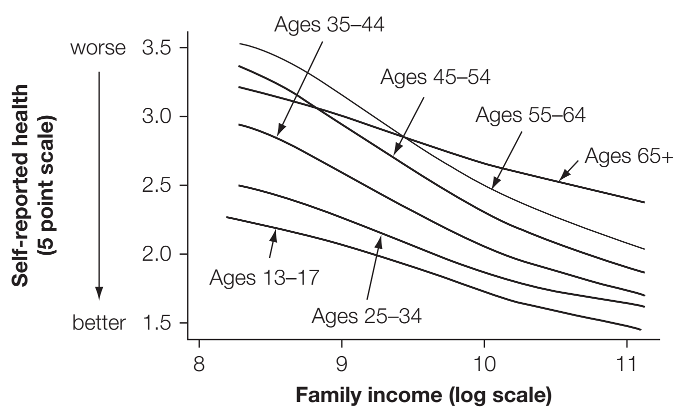


Figure 4.4. *Health inequalities by age.*

Source: From Figure 1 in Case et al. (2002). With permission of the American Economic Association.

axis represent worse self-reported health. The downward slope of each line means that richer people are more likely to report better health.

Table 4.1. *Median wealth by self-reported health status.*

1984 health status	1984 wealth	1994 wealth
Excellent	68.3	127.9
Very good	66.3	90.9
Good	51.8	64.9
Poor	39.2	34.7

Wealth is reported in thousands of 1996 dollars.

Source: Based on Table 1 in Smith (1999). With permission of the American Economic Association.

Self-reported health on a five-point scale may seem a crude measure of health status, but it is more meaningful than it first appears. In a national US study of individuals, study participants ranked their health on this scale in 1984. They were also asked detailed questions about their wealth. Table 4.1 corroborates the result from Figure 4.4: those with better self-reported health tend to be wealthier. Ten years later, the individuals who reported poor health in 1984 had seen a drop in household wealth, while those who reported excellent health in 1984 saw their wealth nearly double. Apparently, self-reported health status is a powerful predictor of future wealth.

Health disparities across income levels exist for more objective measures of health as well. Several different patterns occur, including cases where the poor have worse outcomes, where the rich have worse outcomes, where the disparities emerge only with age, and where no disparities exist at all.

Young man suffering from a hay fever attack – a small price to pay for being wealthy. *Credit: © lichtmeister – Fotolia.com.*

For instance, in the US, children from wealthy families are less likely to suffer from congenital heart defects (see Figure 4.5; in this figure, as well as in Figure 4.4, higher values represent *worse* health). This pattern is consistent with the gradient we have seen so far in mortality and self-reported health. Furthermore, this evidence from newborns indicates that disparities in health may emerge even before birth.

Unlike congenital heart defects, the prevalence of hay fever displays a distinctly different pattern. There is no measured difference between rich and poor among infants, but among older children, the rich are more likely to have been diagnosed with hay fever. One explanation is that children from richer families are more likely to see a doctor and thus more likely to be diagnosed than children from poorer families.

Finally, bronchitis prevalence does not seem to vary consistently with income. As children age, they are more likely to be diagnosed with myopia or have other vision problems, but no difference in the rate of diagnoses emerges between poor and rich. At young ages, the prevalence–income relationship is flat, with poor and rich children equally likely to be diagnosed. In pre-pubescence and adolescence (as with hay fever), richer children are

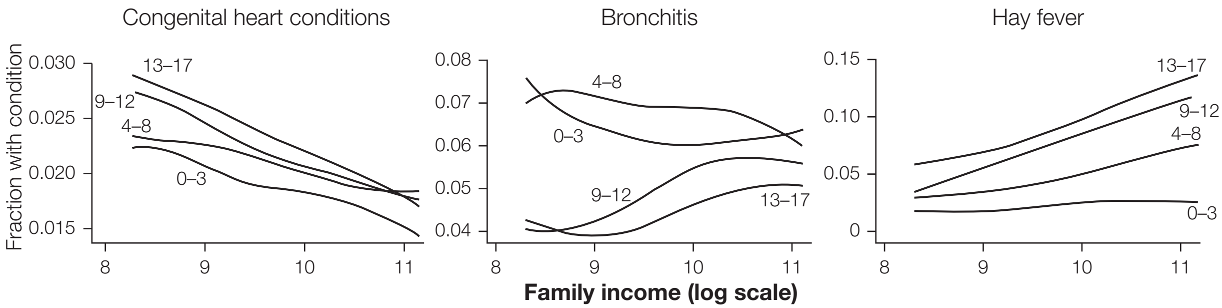


Figure 4.5. *Health inequalities by condition.*

Source: Figure 2 in Case et al. (2002). With permission of the American Economic Association.

more likely to be diagnosed with bronchitis. But hay fever and bronchitis are exceptions to the usual rule that poorer children are more likely to be sick than richer children.

Disparities in countries with universal health insurance

While this evidence of health disparities from the US is overwhelming, we might wonder whether different countries with different health systems have fewer or less obvious socioeconomic disparities. A country like Canada, for instance, has universal health insurance, which may reduce health outcome differences between rich and poor. But the evidence shows health disparities persisting in Canada as well as other countries with different health care systems. Infants born to poor families in Canada are nearly twice as likely to have poor self-reported health than those born to rich families. This gap in self-reported health widens dramatically after age ten (see Figure 4.6).

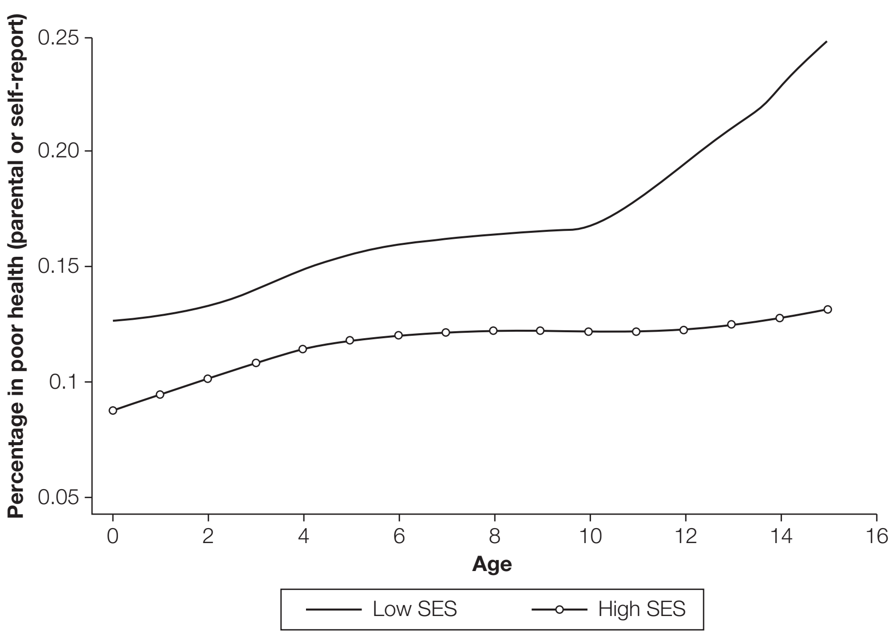


Figure 4.6. *Health inequalities among Canadian children.*

Source: Currie and Stabile (2003). With permission of the American Economic Association.

Disparities across races

Health disparities arise not just between college graduates and high-school graduates or between the rich and the poor. There are also substantial differences in health status and outcomes between different racial groups. Figure 4.7 shows evidence from the US that Hispanic individuals report better health than black individuals, and white individuals report better health than both groups. Health naturally deteriorates with age for all groups, but racial disparities remain.

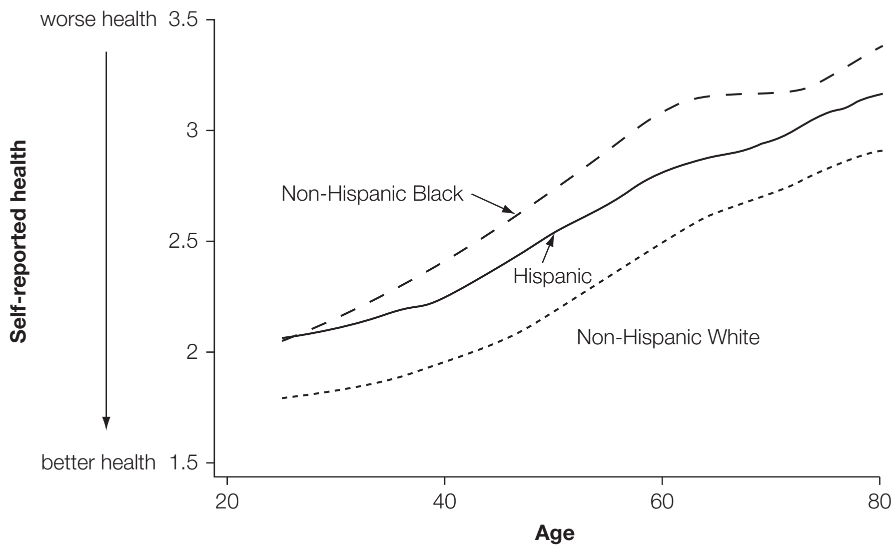


Figure 4.7. *Health inequalities by race.*

Notes: The curves reflect local linear regression estimates. The regressions are weighted using the survey weights provided by the NHIS.

Source: From *The Oxford Handbook of Health Economics*, edited by Sherry Glied and Peter C. Smith (2011), Ch. 7, Socioeconomic status and health: dimensions and mechanisms, by David M. Cutler, Adriana Lleras-Muney, and Tom Vogl, pp. 124–163, Figure 7.5b from p. 134. Reprinted by permission of Oxford University Press.

Disparities in non-human societies

Finally, health disparities between different social groups have been found even in non-human societies. Biologist Robert Sapolsky has spent decades studying social interactions and hierarchies among baboon troops in East Africa. He consistently found that dominant baboons at the top of their social hierarchies are in better health than subordinate baboons. For instance, Sapolsky and Mott (1987) show that dominant baboons have higher levels of high-density lipoprotein (HDL) – the "good cholesterol" – which in humans is correlated with lower rates of heart disease. Sapolsky has argued that this disparity arises because dominant baboons suffer lower stress: they relax by pounding the heads of subordinate baboons.

Summary of evidence for health disparities

Table 4.2 summarizes the evidence we have discussed for health disparities in various contexts. We argue that health disparities are not an accidental or atypical outcome but are a pervasive feature of human (and non-human) societies. It is tempting to try to explain

Table 4.2. *Summary of evidence for health disparities in different populations.*

Factor	Measure of health	Disparity	Evidence
Education	Life expectancy	Better-educated live longer	Male survival curves (Fig. 4.2)
Wealth	Self-reported health	Wealthier report better health	NHIS data from US (Fig. 4.4), Canadian kids (Fig. 4.6)
	Prevalence of congenital heart disease and other conditions	Wealthier less likely to have disease	NHIS data (Fig. 4.5)
Race	Self-reported health	Whites report better health than Hispanics, blacks	NHIS data from US (Fig. 4.7)
Social standing	Life expectancy	Dominant class lives longer	Ducal study in England (Fig. 4.3)
	HDL levels	Dominant class has better HDL levels	Sapolsky study of baboons

each instance of this phenomenon with its own theory. For example, one theory for racial disparities in health is racial discrimination by doctors and nurses. There may be some truth to this theory, but even so it could not possibly explain the evidence for health disparities in industrial-age England or baboon troops in the Serengeti. We search for broader theories that work in multiple settings.

In the next sections, we discuss the most prominent theories that have been advanced to explain health disparities and the evidence supporting each. These theories are not mutually exclusive, and it seems likely that no single theory will ever explain all of the many health disparities.

4.2 The Grossman model and health disparities

The Grossman model discussed in Chapter 3 provides an excellent organizing framework for thinking about health disparities. It encapsulates a family of hypotheses that have been advanced to explain health differences across socioeconomic groups.

Recall from our discussion of the Grossman model that the marginal efficiency of capital (*MEC*) curve plays a critical role in determining optimal health. The *MEC* curve indicates how much lifetime utility each additional unit of health capital creates. If two individuals have different *MEC* curves, as in Figure 4.8, they

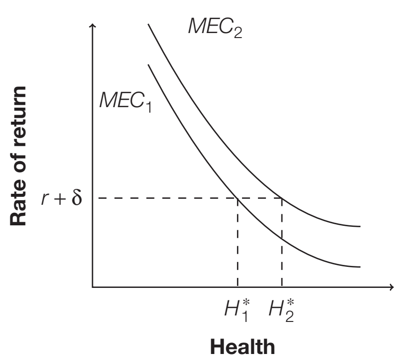


Figure 4.8. *Two* MEC *curves.*

will choose different optimal levels of health even if they are identical in every other way. Each of the hypotheses we discuss in this section implies different *MEC* curves for different socioeconomic groups, and hence different optimal health levels.

How would different *MEC* curves arise for different groups? The Grossman model has many moving parts and parameters, such as health productivity, resource constraints, health depreciation rate, total productive time, and rate of time discounting. Differences in any of these parameters can result in different *MEC* curves and different optimal health decisions. Although not always framed this way by researchers, each of the theories we discuss below can be interpreted in terms of the Grossman model.

4.3 The efficient producer hypothesis

The fact that education and health are correlated does not necessarily imply a causal connection between them. Lleras-Muney (2005) searches for a causal effect of education by statistically exploiting changes in compulsory education laws in the US during the early twentieth century. She finds that an additional year of schooling caused as much as a 1.7-year increase in life expectancy for those born in 1925.

So Lleras-Muney (2005) finds that education improves health, but the mechanism is not clear. The **efficient producer hypothesis** postulates that better-educated individuals are more efficient producers of health than less well-educated individuals. Recall that the Grossman model predicts that people who are more efficient health producers will have higher levels of optimal health. If this hypothesis is accurate, then we expect a health disparity across people with different educational levels, such as the mortality–education gradient found by Lleras-Muney.

Definition 4.1

Efficient producer hypothesis: health disparities exist because better-educated individuals are more efficient producers of health than less well-educated individuals.

There are many reasons why education might lead to more efficient health production. For instance, there may be lessons learned in school that allow students to take better care of themselves. Alternatively, schooling may teach students to be more patient and more willing to make investments that pay off over a long period of time, like health.

A third possible mechanism is that better-educated consumers are more likely to adhere to treatment regimens that require diligence, and better able to navigate complex treatment plans. Consumers who are better readers may also be more able to follow directions that come with prescription drugs. Those more comfortable with math skills may have an easier time figuring out the right dose of insulin, or following through with a diet that requires careful tracking of calories.

Goldman and Smith (2002) focus on the possibility that differences in patient self-management can explain the relationship between health and education. They study people with two diseases, HIV and Type I diabetes, that require intensive patient participation in treatment. HIV patients undergoing anti-retroviral therapies must be vigilant about taking multiple drugs each day without fail; even a few missed doses can lead to a higher level of virus within the body. Diabetics, similarly, must calibrate doses of insulin carefully several times a day in order to control blood sugar levels and reduce the risk of complications.

If there is any validity to the efficient producer hypothesis, then highly educated patients should be better able to adhere to these complex treatment regimens. In fact, this is exactly what the data says. In the study, better-educated HIV patients had better self-reported adherence, defined as having taken all medications on schedule in each of the last seven days. Education, in this case, seems to correlate strongly with the sorts of skills that permit HIV patients to keep track of which drugs they are supposed to take, the appropriate doses, and when they need to take them. This adherence differential explains entirely the health disparity between well-educated and poorly educated HIV patients. Highly educated HIV patients are thus more efficient producers of health.

For diabetic patients, Goldman and Smith found that a randomized intervention promoting intensive adherence to insulin therapy had different effects on people with different educational levels. The less schooling people had, the more effective the intensive treatment was on future adherence as compared with those with more schooling. In the study, patients randomly assigned to the intensive support treatment were given more help to stay on their prescribed therapy. This is exactly as predicted by the efficient producer hypothesis since people with more education presumably do not need this extra support to adhere to their prescribed insulin therapy. See Exercise 12 for more on their results.

4.4 The thrifty phenotype hypothesis

Adherence to medical advice is not the only way that people can differ in their health productivity. There may be genetic reasons why some individuals are less efficient at producing health, even with the same resources. Strictly speaking, this theory is a form of the efficient producer hypothesis, but we devote a separate section to it because of the prominent and distinct role this theory has played in the scholarly debate.

The **thrifty phenotype hypothesis** posits that health outcomes throughout life are determined in part by deprivation that occurs in early childhood or even in the womb. Children from poorer families may suffer more deprivation during gestation and infancy, which may explain observed health disparities.

The thrifty phenotype hypothesis suggests that the link between early deprivation and negative adult health outcomes is a result of gene activation. Children born during times of resource deprivation are more likely to activate certain "thrifty" genes that optimize for sparse conditions, for example ones that instruct cells to hoard fat within the body. These children are well-adapted for famine conditions, but may suffer relatively poor health if they end up living in environments with abundant resources. Their thrifty genes, so helpful for deprivation during infancy, may predispose them to obesity, diabetes, and other undesirable health outcomes in adulthood.

Definition 4.2

Thrifty phenotype hypothesis: resource deprivation *in utero* and during early childhood can lead to activation of "thrifty" genes optimized for sparse conditions. Individuals with such genes activated are poorly adapted for abundant conditions and may develop diabetes, obesity, and other disorders. Health disparities arise because poorer individuals are more likely to face this sort of deprivation early in life.

Researchers also refer to this hypothesis as the **Barker hypothesis**.

The British National Cohort Study of children born in March 1946 provides evidence that adverse health events in early life, plausibly the result of poverty, are linked to poor adult health outcomes. Individuals with low birth weight, slow fetal growth, or respiratory illnesses as infants were more likely to suffer from hypertension, chronic obstructive pulmonary disease, and schizophrenia as adults (Wadsworth and Kuh 1997). This is one of several national longitudinal studies, which track the health of survey participants over time, that show similar trends. For instance, Coneus and Spiess (2012) show similar findings in Germany.

But this study does not definitively establish a causal link between early-life deprivation and poor health outcomes. The children who suffered health problems in early life in the British study could have been different from the other children in several other ways, and perhaps some of those differences drove the adult health outcomes. The best way to establish causality is a scientific experiment like the RAND HIE, but in this case that would involve randomly assigning children to prenatal deprivation. That would allow us to compare a group of children with early-life deprivation (the treatment group) to a substantially similar group of children who enjoyed abundant resources early in life (the control group), but such an experiment would be extremely unethical.

Instead, researchers look to evidence from **natural experiments**, environmental shocks that create a treatment group and a control group naturally. In the last year of World War II, for example, parts of Holland suffered a devastating famine as the result of a German blockade. While tragic, this shock created two groups of Dutch babies: those in gestation during the famine, and those conceived immediately afterwards. The groups were demographically similar, except for the fact that the first group experienced massive nutritional deprivation *in utero*, while the latter group did not experience the same extreme level of deprivation.

The major problem with non-experimental studies is *selection bias*, the bias introduced when treatment is chosen and that people who select into treatment differ in important ways from people in the control group. The virtue of a good natural experiment is that it limits selection bias.

Definition 4.3

Natural experiment: a study that uses an environmental shock that creates a treatment group and a control group naturally. This is useful for identifying causal effects because it eliminates selection bias.

Examples of potential natural experiments include:

- a famine that affects one cohort of babies and not another;
- an earthquake that affects only one half of a country;
- a spike in immigration that affects one region and not another;
- a government policy implemented in one state but not a similar neighboring state;
- an unusually large snowfall one winter but not the next.

Evidence developed by Roseboom et al. (2001) shows that individuals gestating during the famine suffered worse health in adulthood than those not conceived until the famine was over (Table 4.3). The babies exposed to starvation *in utero* grew up to experience higher rates of diabetes, lower levels of HDL (good cholesterol), higher incidence of obstructive airway disease, and worse overall health, although not all these results are statistically significant. According to the thrifty phenotype theory, this is because these individuals' genes were optimized for starvation conditions, and not as well adapted for postwar abundance as their younger counterparts' genes.

Table 4.3. *Adult characteristics according to timing of prenatal exposure to the Dutch famine.*

Cohort	Late gestation during famine	Early gestation during famine	Conceived after famine
Type 2 diabetes	21%	16%	15%
HDL cholesterol (mmol/l)	1.32	1.26*	1.32
Obstructive airway disease	15.0%	23.0%	17.3%
General health poor	6.4	10.3*	5.3

* Indicates statistically significant difference from the unexposed cohort.

Source: Table 1 from Roseboom et al. (2001). Copyright (2001), with permission from Elsevier.

The Dutch famine is not the only famine to be studied as a natural experiment. Between 1958 and 1961, China implemented a set of reform policies known as the Great Leap Forward. The reforms included the collectivization of farms and the reassignment of farmers from agriculture to heavy industry. These in combination with other policies led to a massive famine that killed tens of millions throughout China. Chen and Zhou (2007) compare those *in utero* during the famine against those conceived after the famine. They find that those who lived through the famine *in utero* and survived to adulthood were about three centimeters shorter than those conceived afterwards. These survivors also worked shorter hours and earned less income than their slightly younger counterparts. This latter evidence is also consistent with efficient producer hypothesis.

Economist Douglas Almond and his colleagues have studied three other natural experiments in search of the link between early-life deprivation and adult outcomes: the 1918 influenza epidemic (Almond 2006); the 1986 Chernobyl disaster (Almond et al. 2009); and Ramadan, the Muslim holy month of fasting (Almond and Mazumder 2007). In each case, researchers compared individuals affected by the events *in utero* with others who were conceived immediately thereafter, or others who lived in nearby unaffected areas in the case of the Chernobyl incident. In each case, the evidence supports the hypothesis that shocks in early life have adverse impacts on outcomes in later life.

These findings have spurred interest in policy changes that improve fetal health. For example, Currie and Walker (2011) study the introduction of the electronic toll collection system E-ZPass in New Jersey around 1999. The E-ZPass system streamlined payments at toll plazas, eliminating long lines of idling cars and sharply reducing air pollution in the surrounding environment. The authors study babies born to women living within two kilometers of a toll plaza, and found the risk of birth prematurity fell by 10.8% and the likelihood of low birth weight dropped by 11.8% after the introduction of E-ZPass. The thrifty phenotype hypothesis suggests that these babies will enjoy better health outcomes throughout life thanks to less pollution exposure *in utero*.

4.5 The direct income hypothesis

In some sense, the evidence that the rich are in better health is utterly unsurprising. Richer people are better off on many dimensions: by virtue of their wealth, they can afford to send their children to better schools, drive fancier cars, or dine at nicer restaurants. Similarly, they can afford better doctors, gym memberships, and live in healthier neighborhoods with more outdoor parks and less pollution. From this viewpoint, there is nothing mysterious to explain about health disparities.

Definition 4.4

Direct income hypothesis: health disparities arise because the rich have more resources available to invest in health.

The key to better health, according to Lindahl (2005).
Credit: © philhol – Fotolia.com.

Again, the Grossman model provides a way to think about this argument. Consider two individuals who are identical in every way except that individual R earns a high wage, and individual P earns a low wage.

As Figure 4.9 illustrates, the rich individual R has an expanded production possibility frontier (PPF) because of the extra financial resources at her disposal. Her optimal levels of health H and home good Z are both higher than the optimal levels of her poorer counterpart. Staying healthy is more important to the richer individual because her time is more valuable. Because the returns to health are greater, her MEC curve will be shifted up relative to the MEC curve of the poorer individual.

Differences in wage are not the only way that financial resources can create health disparities. People with more wealth, whatever the source, have more resources available for health production. So the Grossman model would predict better health even for lottery winners, even though their productivity and wages do not change overnight as a result of their great luck. Lindahl (2005) shows that this is indeed the case: a

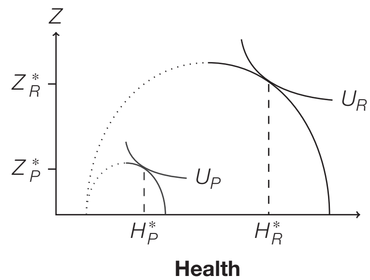


Figure 4.9. *The rich individual* R *earns higher wages than the poor individual* P *and therefore has more money at her disposal to purchase medical goods and home goods. Consequently, her* PPF *curve is shifted outward and her optimal level of health is higher.*

lottery win that increased income by just 10% decreased the five-year mortality rate from 6% to 4% in a sample of Swedish lottery participants.

4.6 The allostatic load hypothesis

The allostatic load hypothesis of health disparities emphasizes stress as the main mechanism linking socioeconomic status and health. The stress response in humans is vitally important to survival. In the face of threats, the body releases a hormone called adrenaline produced by the adrenal glands located just above the kidneys. This is known as the "fight or flight" response. In response to the adrenaline rush, the body shunts blood to the muscles and lungs, and away from the kidneys and stomach, in an attempt to prioritize the body parts most useful for fighting or fleeing (Sapolsky 1995).

This response can save the life of someone face to face with a saber-tooth tiger. But if the stress response is prolonged or repeated, it becomes unhealthy. The body adapts to frequent stress response by producing hormones like glucocorticoids. In the short run, glucocorticoids shut down the immune system and increase metabolism. But repeated or prolonged exposure to glucocorticoids triggers a biochemical cascade that can lead to memory loss, strokes, and neuron death; it accelerates the aging of the brain. This is one reason why students often get sick after taking final exams: the "fight or flight" response carries them through studying and test-taking but the weakened immune system eventually leads to infection.

In the Grossman model, the aging process is captured in the rate of depreciation of health capital δ. A person under prolonged or repeated stress has a higher δ than a person who lives a more relaxed life, even at the same biological age. Figure 4.10 shows the consequences of high stress: optimal health will be lower because health investment is less worthwhile. This holds even though the efficiency of health production is the same (that is, even though both individuals face the same *MEC* curve).

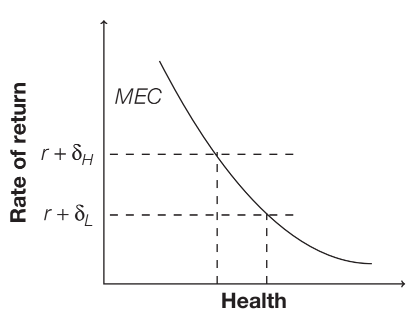


Figure 4.10. *An individual with prolonged or repeated stress faces a higher rate of health depreciation* δ_H *and chooses a lower optimal health level as a result.*

In modern life, the stress response is more frequently triggered by threats that look nothing like a saber-tooth tiger, such as final exams, overbearing bosses, and mortgage payments. Such repeated stress creates a cumulative physiological burden known as **allostatic load**. The allostatic load theory predicts that people on the lower end of the socioeconomic status suffer more stress and face worse health outcomes as a result.

Definition 4.5

Allostatic load: cumulative physiologic toll exacted on the body over time by efforts to adapt to life experiences.

Source: Seeman et al. (1997).

Marmot et al. (1978) and Marmot et al. (1991) track the health status of British civil servants from the beginning of their career to test the allostatic load hypothesis. These are known as the Whitehall studies, named for the London street where many British government employees work.

The British civil servants are a particularly interesting group for study, since though there are distinct employment grades, people who apply and become civil servants are relatively homogeneous in backgrounds and workplace environments. In addition, all British civil servants are enrolled in the National Health Service and have similar access to health care. Given this homogeneity, sharp gradients in health disparity between high-grade and low-grade employees would be surprising.

The Whitehall studies find that morbidity and mortality rates are highest among the lowest-grade civil servants, and conversely, lowest among the highest-grade civil servants. The mortality rate for the lowest-level workers was over three times higher than for the highest-grade administrators. There were also sharp socioeconomic gradients in the rates of heart disease, diabetes, and asthma. Behavioral and physical differences like smoking, obesity, high blood pressure, and high cholesterol explain only part of the disparity.

Nelson's Column in Trafalgar Square, near Whitehall in central London. Despite an exalted socioeconomic status, Admiral Horatio Nelson died at a young age. Credit: Jonathan Ling – iStockphoto.com.

Marmot et al. (1991) find that lower-grade civil servants report lower job satisfaction, lower job control, and more stressful life events. They argue that these civil servants suffer higher allostatic load than their higher-grade colleagues. Consistent with the allostatic load hypothesis, higher job stress correlates strongly with poorer health outcomes. Another strong conclusion from the Whitehall studies is that equalizing access to care does not eliminate health disparities. This echoes the conclusions drawn from the RAND HIE (see Chapter 2).

Smith (1999) makes an explicit comparison between the British civil servants in the Whitehall studies and the baboon troops studied by Sapolsky and Mott (1987). Both high-grade British civil servants and dominant baboons have higher levels of high-density lipoprotein (HDL) cholesterol. Higher levels of HDL cholesterol are correlated with lower rates of heart disease, at least in humans. Figure 4.11 shows HDL cholesterol rates for both British civil servants and baboons. This is suggestive evidence that being on the low end of a social hierarchy produces high allostatic load and adverse physiological outcomes.

The income inequality hypothesis

Perhaps it is not absolute income that determines health outcomes but instead the *distribution of income* within a society. Societies with more unequal income distributions,

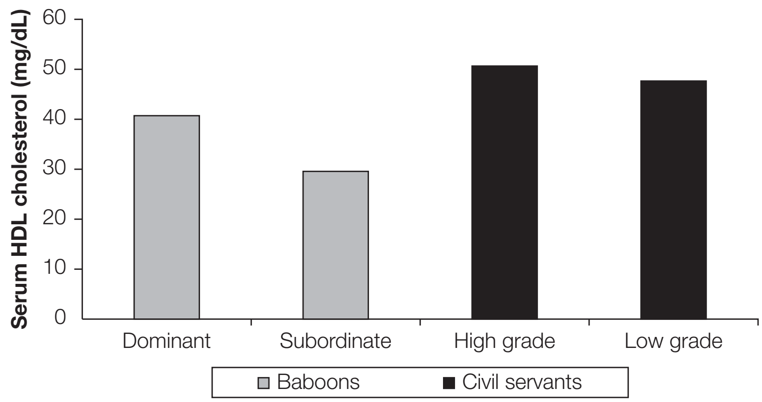


Figure 4.11. *HDL cholesterol levels in baboons and British civil servants by social status.*

Source: Figure 1 from Smith (1999). Reprinted with permission from the American Economic Association.

according to **income inequality hypothesis**, will have worse health. Under this theory, more equal societies are also less stressful ones, and therefore, according to the allostatic load hypothesis, are generally more healthy (Wilkinson and Pickett 2006).

Definition	4.6

Income inequality hypothesis: health disparities are caused by income inequality, which in and of itself is a source of allostatic load for the poor.

The income inequality hypothesis, if true, has important policy implications. If in fact higher levels of inequality have a negative impact on health outcomes, then policymakers should aim at reducing inequality within a community, not just elevating average health. The hypothesis also means that health status in a society may decline even as average income rises if that increase is confined just to the richest groups.

There is a vast empirical literature attempting to document the correlation between society-wide inequality and average health status (Wilkinson and Pickett 2006). All of these studies rely on aggregate data at the international level, the state level, or the country level, and are vulnerable to omitted variable bias. There is much disagreement in the literature about whether the correlation between societal inequality and health is causal (Deaton 2003).

Access to care hypothesis

One advantage of high income is the ability to afford more generous health insurance and more lavish health care. Conversely, people with lower income tend to have less generous insurance and face higher prices for care. This reasoning leads to another possible explanation for health disparities: differences in access to health care.

Bindman et al. (1995) find that low self-reported access to care is strongly predictive of higher rates of hospitalization for chronic conditions: asthma, hypertension, congestive heart failure, and chronic obstructive pulmonary disease. These hospitalizations are largely preventable and are considered indicators of poor overall health and inadequate outpatient care. As expected, it is the uninsured or Medicaid-insured who are more likely to report that it is "very difficult" or "extremely difficult" to get medical care when needed, perhaps because these lower-income individuals do not have the time to seek care or the money to pay for it.

The Oregon Medicaid study, covered in Chapter 2, found similar evidence when examining people randomly assigned to Medicaid coverage. Though mortality rates did not differ due to insurance coverage, diabetic patients with Medicaid were more likely to be diagnosed and to be taking appropriate medication. One possible mechanism is better access to care: those on Medicaid were more likely to report having a regular place of care and a personal doctor (Baiker et al. 2013; Finkelstein et al. 2011).

While a plausible source of health disparities, access cannot explain the whole story. The fact that health disparities persist even in societies with universal health insurance coverage suggests that equalizing access to care does not eliminate health disparities. Two studies we have already discussed – the study of Canadian youth (Currie and Stabile 2003) and the Whitehall studies (Marmot et al. 1978, 1991) – illustrate this point. In the Canadian study, health inequalities between the rich and the poor expanded during adolescence, despite universal access to health care in Canada (Figure 4.6). In the Whitehall study, there were sharp health disparities across different job grades, even though the study participants comprised a relatively homogeneous population – British civil servants – with universal access to health care through the National Health Service.

4.7 The productive time hypothesis

So far our discussion of health disparities has focused on the causal pathway leading from socioeconomic status to health. While there is much compelling evidence for this pathway, there is also evidence of the reverse pathway: changes in health can affect subsequent socioeconomic status as well. Indeed, the Grossman model predicts that worsening health diminishes productive time and hence the ability to produce income. We call this the **productive time hypothesis**.

Definition 4.7

Productive time hypothesis: SES differences are caused by disparities in health. Worsening health diminishes productive time and hence the ability to produce income.

Table 4.1 from Smith (1999) shows one indirect piece of evidence that is consistent with the productive time hypothesis. Survey respondents reporting poor health in 1984 lost wealth over the next decade, while those who reported excellent health in 1984 saw their assets nearly double over the same time period.

Smith (1999) also reports more direct evidence from the Health and Retirement Study (HRS). People newly diagnosed in 1992 with severe chronic diseases such as cancer or heart disease had lost on average $17,000 in wealth by 1994, representing about 7% of household income. This loss cannot be explained by higher medical bills alone: over the same period, average out-of-pocket medical expenses totalled only about $2,300.

The respondents who were newly diagnosed with severe chronic conditions in 1992 also tended to work less or leave work completely; 21% left the workforce between 1992 and

1994 in the wake of their diagnosis. Even those who remained employed worked about four fewer hours per week on average. In total, this reduction in working time resulted in a decrease in earnings of about $2,600. This labor force exit is exactly as the productive time hypothesis predicts, although the decrease in earnings is not enough to explain the total decrease in wealth either.

It seems clear from the HRS data that health shocks late in life can alter socioeconomic status. There is also good evidence that health shocks early in life – even before birth – play a fundamental role in determining education levels, poverty, and other economic outcomes.

For instance, Barreca (2010) studies the effect of malaria exposure *in utero* in the US between 1900 and 1936. Babies born in states and years in which temperatures were ideal for *Anopheles* mosquito breeding were substantially more likely to end up in poverty by 1960 than babies born in other states and times. They also tended to end their schooling earlier in life. This evidence is consistent with the productive time hypothesis – the babies exposed to malaria were presumably sicker during childhood and had less time and energy to invest in their own human capital development.

Similarly, Oreopoulos et al. (2008) and Black et al. (2007) study twins and siblings to measure the effect of poor health in infancy on later-life outcomes. The benefit of studying twins and siblings is that they share a common background and typically grow up in similar environments. This research strategy helps isolate the effect of birth weight and other measures of newborn health. Both studies find lower mortality rates for healthier babies, along with higher educational attainment and higher earnings in adulthood. In addition, they find that low birth weight children suffer higher mortality rates relative to their twins or siblings; this is causal evidence in favor of the thrifty phenotype hypothesis.

The Grossman model predicts that the productive time hypothesis is more relevant to people with poor health. Figure 4.12 shows the diminishing marginal returns to health in terms of productive time. Health changes have a larger effect on total productive time for relatively unhealthy individuals like those at point *A* in the figure. Healthier individuals like those at point *B* are in the flat portion of the curve; for them, changes in health have only minor effects on total productive time.

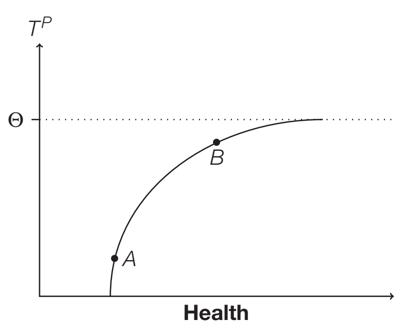


Figure 4.12. *The production function of productive time* T^P.

One consequence of this reasoning is that we should find strong evidence for the productive time hypothesis in poorer countries where the average level of health is lower. Thomas et al. (2004) conducted a randomized experiment in Indonesia in which half the subjects received a weekly iron supplement of 120 mg for a year. The other half took a placebo pill over the same time period.

Male and female subjects who received the iron treatment were much less likely to be anemic. The men who received the treatment were also more likely to work. And among self-employed males, earnings were likely to be much higher. In general, the men receiving iron supplements were more productive than the placebo group. The evidence from this randomized study is consistent with the idea that the productive time hypothesis applies strongly in poorer countries. If this result can be generalized beyond iron supplements and anemia, then small investments in health improvement in poor countries may yield large economic returns.

The iron supplement study is a good example of the productive time hypothesis because of the medical consequences of anemia. Anemia is a condition defined as a shortage of hemoglobin or red blood cells, which carry oxygen throughout the body. Iron deficiency is a common cause of anemia. Because anemia can lead to lethargy and the inability to concentrate, anemic individuals typically have much lower capacity for work. In Grossman parlance, anemic individuals have lower H and greater T^S.

4.8 Time preference: the Fuchs hypothesis

Fuchs (1982) suggests that the observed correlation between socioeconomic status and health is actually caused by an unobserved third variable: patience. People who are patient show high tolerance for delayed gratification. And people who are willing to delay gratification invest more in both education and health status. In other words, patience may explain the correlation between health and education.

Definition 4.8

Fuchs hypothesis: both health and SES disparities are simultaneously caused by innate differences in the willingness to delay gratification. Individuals with a lower rate of time discounting (who are therefore more patient) invest more in both health and education.

The Grossman model provides a good framework for studying this argument. An individual's lifetime utility is

$$\begin{aligned} U &= U(H_0, Z_0) + \delta U(H_1, Z_1) + \delta^2 U(H_2, Z_2) + \cdots + \delta^{\Omega} U(H_{\Omega}, Z_{\Omega}) \\ &= \sum_{t=0}^{\Omega} \delta^t U(H_t, Z_t) \end{aligned} \tag{4.1}$$

where $\delta \in [0,1]$ is the individual's time discount factor (not to be confused with γ, her depreciation rate). When an individual's δ is low, near 0, she is impatient. Utility from future periods is discounted heavily, and even high levels of utility in the far future do not appeal greatly to the individual. When an individual's δ is near 1, by contrast, she is patient; the prospect of high utility in the future is appealing and she will be more reluctant to trade future utility away for utility in the present.

Now consider two individuals, Bart and Lisa, who are identical in every way except that Bart is impatient (with low δ) and Lisa is patient (with high δ). Who has a higher optimal level of health H^*?

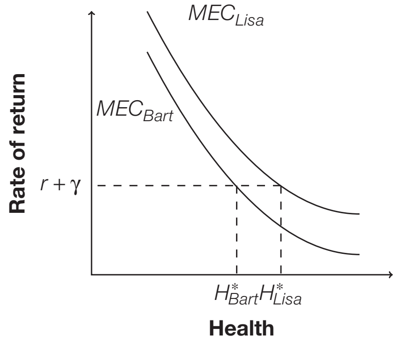


Figure 4.13. *Lisa gains more lifetime utility from a small increase in health than the impatient Bart because patient Lisa values utility in future periods highly. Consequently, her* MEC *curve is shifted upward and her optimal level of health is higher.*

Consider the impact of time discounting on the *MEC* curve. Lisa, who values utility in future periods more, benefits more from each unit of health investment than Bart does because the benefits of health are realized mostly in future periods. Consequently, Lisa's *MEC* curve lies above Bart's *MEC* curve, and her optimal level of health is higher (see Figure 4.13).

This link between patience and investment should be intuitive. To invest in health (or any type of capital), an individual spends resources now in order to gain in the future. For someone who discounts future utility heavily, this is an unattractive option: Why spend time and money to build up health (H) for future periods when it could be spent on gaining utility from other goods (Z) today?

An identical argument can be made with regard to education, which is a different kind of human capital that also requires investment. Students must forgo income, sleeping in, and entertainment in exchange for learning, which pays off only in the future. People who discount the future heavily are less willing to make this tradeoff.

This time-discounting theory predicts that people who do not discount the future very heavily will be well educated and healthy, while people who *do* discount the future heavily will be poorly educated, unhealthy, and the happier for it. Then, even if there is no causal connection between education levels and health, we would still expect to see the strong correlation between education and health that we do observe.

More evidence that suggests that time discounting is an innate quantity rather than a learned quantity comes from research on smoking patterns. Suppose the lessons learned in college cause people to avoid smoking. Then a 17-year-old in twelfth grade would have the same propensity to smoke as any other 17-year-old in twelfth grade, even if one were an aspiring astrophysicist destined for college and the other were planning to enter the workforce right after high-school graduation. Only once the aspiring astrophysicist actually gains an educational advantage will she be less likely to smoke than her high-school peer.

Alternatively, suppose that the Fuchs hypothesis is true. Then the aspiring astrophysicist invests more in education because she innately has more patience than her high-school peers. Under this hypothesis, we would expect the 17-year-old aspiring astrophysicist to *already* be less likely to smoke, even in the twelfth grade. It is not the lessons of college that cause her not to smoke, but instead her innate patience.

Using data on smoking patterns and education history from the Stanford Heart Disease Prevention Program, Farrell and Fuchs (1982) test the Fuchs hypothesis. They find that, even at age 17, students who go on to college have a lower rate of smoking. Hence, future education disparities predict current disparities in smoking rates. This result suggests that time discounting rather than education drives the correlation between education and health.

In a charming demonstration of the persistence of time-discounting attitudes, a different set of Stanford researchers placed four-year-old test subjects in a room alone in front of a plate with a single marshmallow (Mischel et al. 1972). The four-year-olds were told that if they could resist eating the marshmallow for a few minutes, they would be rewarded with several marshmallows at the end of their isolation. While some kids ate the marshmallow as soon as the investigator left the room, others successfully resisted the temptation and were rewarded for their patience.

Following up on this experiment, Shoda et al. (1990) found that the kids who were able to delay gratification as four-year-olds, scored much better on math and language tests as 18-year-olds. The ability to delay gratification is persistent and explains why the kids who waited for the extra marshmallows were also more successful academically, since academic success also requires delayed gratification.

Cutler and Lleras-Muney (2010) find in national survey data from the US and the UK that higher-educated people are less likely to take unnecessary risks like not wearing seat belts and more likely to take preventative health measures like cancer screenings. This result, in itself, is consistent with the Fuchs hypothesis. However, Cutler and Lleras-Muney (2010) find that including proxy variables for time preference *increases* the strength of the relationship between education and risk-taking, contrary to the Fuchs hypothesis. This finding would suggest that heterogeneity in time-discounting rates is not a primary driver of health disparities.

4.9 Conclusion

Each of the theories outlined above works within the logic of the Grossman model, but each one explains the connection between health and socioeconomic status in its own way. The first four theories we discussed – the *efficient producer hypothesis*, the *thrifty phenotype hypothesis*, the *direct income effect* and the *allostatic load hypothesis* – all describe ways in which more wealth or better education lead to improved health. The *productive time hypothesis* reverses that causal argument; it posits that improved health leads to better SES. Finally, the *Fuchs hypothesis* argues that a third variable, time discounting or “patience”, determines both health and wealth.

Each theory has supporting evidence, and it could well be the case that they all help explain socioeconomic health disparities. Figure 4.14 summarizes the causal structure of the Grossman family of theories.

This is an active and contentious area of research, and it seems likely that future work in health economics will clarify issues that at this point remain murky. But a few things have been well established:

- better-educated people are more efficient producers of health, even with the same resources;
- health events early in life have important consequences throughout the lifespan;
- stress plays an important role in producing health disparities;
- equalizing access to care does not eliminate health disparities;
- there is a bi-directional relationship between health and socioeconomic status.

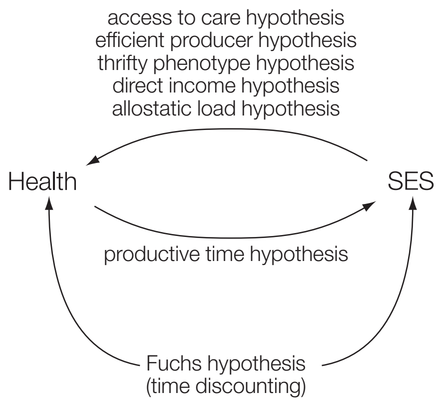


Figure 4.14. *Causal relationship between health, socioeconomic status, and other variables.*

Policies aimed at reducing health disparities must cope with these facts. Because health disparities have a multifaceted set of causes, policymakers must be mindful of unintended consequences when addressing the root causes of health disparities.

4.10 Exercises

Comprehension questions

Indicate whether the statement is true or false, and justify your answer. Be sure to cite evidence from the chapter and state any additional assumptions you may need.

1 In the US, well-educated males can expect to live longer than poorly educated males.

2 Unlike in the US, there are no socioeconomic status gradients in health in countries that provide universal health care coverage to all citizens. That is, in such countries, poorer and richer citizens have (on average) the same health.

3 Health status earlier in life is a good predictor of wealth later in life.

4 According to Smith (1999), nearly all of the differences in health outcomes between rich and poor in America can be attributed to differences in access to medical care.

5 The thrifty phenotype hypothesis states that early-life events after birth have a strong influence on health status even in adulthood.

6 People who have a newly diagnosed chronic disease, such as diabetes, often suffer large declines in their wealth over time. This decline in wealth is entirely explained by decreased hours of work.

7 In the Whitehall study, access to health care was a key variable determining the relative health outcomes of high- and low-grade British civil servants.

8 One leading theory about why the poor are in worse health than the rich is that the rich enjoy a greater allostatic load.

9 In a study of babies born during the Dutch famine toward the end of World War II, those exposed to the famine *in utero* were more likely than those not as exposed to be obese as adults.

10 In Canada, unlike in the US, the gap between rich children and poor children in health status does not widen as children age.

11 There is a consensus among health economists that socioeconomic status has a major impact on health, but health does not have a significant effect on SES.

Analytical problems

12 Table 4.4 shows results from the study of diabetic patients discussed in Goldman and Smith (2002). Recall that three groups of patients separated by education level were treated with one of two therapies (conventional vs. intensive). Their average hemoglobin A1c levels (also known as glycosylated hemoglobin) before and after the experiment are reported. Note that, as in golf, lower scores are better on the hemoglobin A1c test, which measures how well the patient has kept blood glucose levels in control. Most treatment regimens for diabetes are designed to keep hemoglobin A1c levels low.

Table 4.4. *Improvement in hemoglobin measurement for diabetics by treatment regime and age.*

Group	Glycosylated hemoglobin		
	Postgraduate degree	College grad/ some college	HS degree/some secondary
Conventional therapy only ($n = 495$)			
Baseline	8.42	8.76	8.96
End-to-study	8.88	9.08	9.59
Difference	0.46	0.32	0.63
Intensive treatment only ($n = 490$)			
Baseline	8.04	8.86	8.93
End-of-study	7.18	7.30	7.43
Difference	−0.85	−1.56	−1.51
Treatment effect*	−1.31	−1.88†	−2.14‡

* Treatment effect is the improvement in glycemic control among the intensive treatment group relative to conventional therapy. Average follow-up period was 72 months. Significance levels are for a test of equivalence with the postgraduate category and control for duration in study, gender, marital status, and age. Intensive treatment was more efficacious for the less educated.
† $P < 0.10$.
‡ $P < 0.05$.

Source: Goldman and Smith (2002). Copyright (2002) National Academy of Sciences, USA. Reproduced with permission.

As the figure indicates, average hemoglobin A1c levels deteriorated for patients in the control group, but improved for patients in the treatment group.

a In all three groups of conventional-therapy patients, glucose levels got worse. Which of the conventional therapy groups had the biggest deterioration in average glycosylated hemoglobin levels?

b Which intensive-treatment group had the biggest improvement in average glycosylated hemoglobin levels?

c For which group of patients did the intensive-treatment intervention seem to have the biggest effect compared with the conventional therapy? The smallest effect?

d Explain how this evidence is consistent with the efficient producer hypothesis.

13 Table 4.5 shows data from Rich-Edwards et al. (2005) on the prevalence of various afflictions among female nurses who were born at different weights.

Table 4.5. *Hazard rate of coronary heart disease, stroke, and total cardiovascular disease, compared with average birth rate cohort.*

Cohort	Average birth weight	Low birth weight	Very low birth weight	Extremely low birth weight
Birth weight (kg)	3.2–3.9	2.5–3.2	2.3–2.5	< 2.3
Coronary heart disease	100%	130%*	148%*	131%
Stroke	100%	116%	105%	123%
All cardiovascular disease	100%	123%*	129%*	127%

* Indicates statistically significant difference from the average birth weight cohort.
Hazard ratios are adjusted for age and body mass index (BMI).

Source: Data from Table 1 in Rich-Edwards et al. (2005).

- **a** Is it a coincidence that all nurses born at average birth weight suffer from coronary heart disease, stroke, and all cardiovascular disease at the exact same rate?
- **b** Summarize the data in the figure in one or two concise sentences.
- **c** Discuss the data in light of what you know about the thrifty phenotype hypothesis.

14 Simpson's paradox. Recently, the king of the mythical nation of Pcoria was diagnosed with cancer. He noticed that many other nobles in his court had also been diagnosed with cancer recently. He dispatched his two best health economists to study cancer rates in Pcoria's two towns. Their results are recorded in Table 4.6. Before you begin, recall that Pcoria's two towns are very different. Eastville is a poor area, home to numerous factories that produce prodigious amounts of pollution. Weston is the home of the king's court and is filled with pristine parks and hiking trails.

Table 4.6. *Information from the cancer rate study commissioned by the King of Pcoria.*

	Eastville		Weston	
	Population	Cancer rate	Population	Cancer rate
Nobles	100	50%	900	10%
Peasants	1,000	40%	500	8%

- **a** The king's economists conclude that there is a health disparity between nobles and peasants when it comes to cancer rates. Succinctly state the health disparity.
- **b** Calculate the nationwide cancer rate for nobles and the nationwide cancer rate for peasants. What do you find?
- **c** This effect is called *Simpson's paradox* or the amalgamation paradox. Explain how the distribution of nobles and peasants and the geography of Pcoria contribute to the paradoxical result. Which health disparity hypothesis from the chapter does this exemplify?
- **d** Consider Table 4.7, which substitutes variable names for the cancer rate data. Simpson's paradox occurs when the cancer rate is higher for nobles in each town but higher for peasants nationwide. Express these three conditions in terms of the variable names in the table.
- **e** Use your answer from Exercise d to prove that Simpson's paradox cannot occur if the populations of nobles and peasants are the same in both towns (that is, $a = c$ and $b = d$).

Table 4.7. *Generalized version of Table 4.6.*

	Eastville		Weston	
	Population	Cancer rate	Population	Cancer rate
Nobles	a	w	c	y
Peasants	b	x	d	z

f Do you think Simpson's paradox might invalidate some of the evidence for health disparities cited in this chapter? Why or why not?

Essay questions

15 The following is an excerpt from the abstract of a recent journal article entitled "The effects of maternal fasting during Ramadan on birth and adult outcomes" by Almond and Mazumder (2007). Ramadan is the traditional month of daytime fasting by Muslims.

> We use the Islamic holy month of Ramadan as a natural experiment for evaluating the short and long-term effects of fasting during pregnancy. Using Michigan natality data we show that *in utero* exposure to Ramadan among Arab births results in lower birthweight and reduced gestation length. Preconception exposure to Ramadan is also associated with fewer male births. Using Census data in Uganda we also find that Muslims who were born nine months after Ramadan are 22 percent ($p = 0.02$) more likely to be disabled as adults. Effects are found for vision, hearing, and especially for mental (or learning) disabilities.

a Describe how one or more theories discussed in this chapter might explain the findings by Almond and Mazumder.

b Suppose that a scientific study determines that fasting during Ramadan actually has no causal effect on fetal health. What other factor could explain the Michigan results? Do you think your explanation is likely?

16 Below is the abstract of a recent National Bureau of Economic Research working paper entitled "Positive and negative mental health consequences of early childhood television watching" by Waldman et al. (2012):

> An extensive literature in medicine investigates the health consequences of early childhood television watching. However, this literature does not address the issue of reverse causation, i.e., does early childhood television watching cause specific health outcomes or do children more likely to have these health outcomes watch more television? This paper uses a natural experiment to investigate the health consequences of early childhood television watching and so is not subject to questions concerning reverse causation. Specifically, we use repeated cross-sectional data from 1972 through 1992 on county-level mental retardation rates, county-level autism rates, and county-level children's cable-television subscription rates to investigate how early childhood television watching affects the prevalence of mental retardation and autism. We find a strong negative correlation between average county-level cable subscription rates when a birth cohort is below three and subsequent mental retardation diagnosis rates, but a strong positive correlation between the same cable

subscription rates and subsequent autism diagnosis rates. Our results thus suggest that early childhood television watching has important positive and negative health consequences.

- **a** Assuming the findings of this paper are correct, connect their results to one or more of the hypotheses discussed in this chapter.
- **b** The researchers assume that higher rainfall in a county will lead to higher cable subscription rates (because people in those counties are more likely to stay inside and watch TV). In other words, they treat rainfall amounts as a random influence on cable subscription rates. Explain why a natural experiment is important to identify a causal effect. In this case, what selection biases are the researchers concerned about?
- **c** Assume that this natural experiment is invalid and that cable subscription rates are instead totally determined by income (that is, rich families are more likely to buy cable, and the local weather plays no role). Interpret their evidence in light of this possibility.

Students can find answers to the comprehension questions and lecturers can access an Instructor Manual with guideline answers to the analytical problems and essay questions at **www.palgrave.com/economics/bht**.

This page intentionally left blank